

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2005, 09:42:08 ; Search time 131.5 Seconds  
(without alignments)  
16.706 Million cell updates/sec

Title: SEQ-RGRGR  
Perfect score: 27  
Sequence: 1 rgrgr 5

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 1000 summaries

Database : A\_Geneseq\_21.\*  
1: Geneseq1980s.\*  
2: Geneseq1990s.\*  
3: Geneseq2000s.\*  
4: Geneseq2001s.\*  
5: Geneseq2002s.\*  
6: Geneseq2003as.\*  
7: Geneseq2003bs.\*  
8: Geneseq2004s.\*  
9: Geneseq2005s.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	100.0	7	2	AAW65560
2	27	100.0	7	4	AAE09064
3	27	100.0	7	6	ABU07629
4	27	100.0	7	8	ADI20905
5	27	100.0	7	8	ADOI7054
6	27	100.0	8	2	AAW65559
7	27	100.0	8	4	AAE09063
8	27	100.0	8	6	ABG75621
9	27	100.0	8	6	ABB82936
10	27	100.0	8	6	ABU07628
11	27	100.0	8	8	ADOI7084
12	27	100.0	8	8	ADOI7088
13	27	100.0	9	2	AAW60714
14	27	100.0	9	2	ADR21207
15	27	100.0	9	8	ADS18175
16	27	100.0	10	4	AAG95106
17	27	100.0	10	4	AAG94476
18	27	100.0	10	4	AAG95290
19	27	100.0	10	4	AAG97172
20	27	100.0	10	4	AAG95526
21	27	100.0	10	4	AAG97414
22	27	100.0	10	4	AAG97184
23	27	100.0	10	4	AAG95388
24	27	100.0	10	4	AAG97174

25	27	100.0	10	4	AAG94474
26	27	100.0	10	4	AAG94478
27	27	100.0	10	4	AAG95174
28	27	100.0	10	4	AAG95528
29	27	100.0	10	4	AAG97180
30	27	100.0	10	4	AAG94834
31	27	100.0	10	4	AAG95878
32	27	100.0	10	4	AAG96208
33	27	100.0	10	8	ADJ66128
34	27	100.0	11	2	AAW09224
35	27	100.0	11	2	AAV01039
36	27	100.0	11	3	AAV51894
37	27	100.0	11	3	AAV90997
38	27	100.0	11	5	ABG31434
39	27	100.0	11	7	ABW01640
40	27	100.0	11	8	ADJ96780
41	27	100.0	11	8	ADRA5935
42	27	100.0	11	8	ADRA5929
43	27	100.0	11	8	ADRA5932
44	27	100.0	12	2	AAW65568
45	27	100.0	12	2	AAV01044
46	27	100.0	12	4	AAE09159
47	27	100.0	12	6	ABU07643
48	27	100.0	12	8	ADOI7059
49	27	100.0	13	6	ABB82931
50	27	100.0	13	6	ABB82932
51	27	100.0	14	5	ABP54101
52	27	100.0	14	6	ABB82933
53	27	100.0	14	6	ABB82934
54	27	100.0	14	6	ABB82935
55	27	100.0	14	7	ADC19846
56	27	100.0	14	8	ADS52349
57	27	100.0	15	1	AAE06481
58	27	100.0	15	2	AAE25058
59	27	100.0	15	4	ABP25029
60	27	100.0	15	5	ABP51969
61	27	100.0	15	7	ADF89480
62	27	100.0	15	7	ADF89481
63	27	100.0	15	8	ADK48862
64	27	100.0	15	8	ADK48863
65	27	100.0	15	8	ADS52340
66	27	100.0	15	8	ADS52345
67	27	100.0	15	8	ADS52330
68	27	100.0	15	8	ADS52335
69	27	100.0	15	8	ADS52343
70	27	100.0	15	8	ADS52337
71	27	100.0	15	8	ADS52338
72	27	100.0	15	8	ADS52341
73	27	100.0	15	8	ADS52334
74	27	100.0	15	8	ADS52331
75	27	100.0	15	8	ADS52342
76	27	100.0	15	8	ADS52332
77	27	100.0	15	8	ADS52339
78	27	100.0	15	8	ADS52333
79	27	100.0	15	8	ADS52336
80	27	100.0	16	2	AAW65569
81	27	100.0	16	4	AAE01042
82	27	100.0	16	4	AAE09160
83	27	100.0	16	4	AAG64369
84	27	100.0	16	6	ABU07644
85	27	100.0	16	8	ADOI7080
86	27	100.0	17	9	AAE01837
87	27	100.0	19	2	AAV02639
88	27	100.0	19	2	AAV01038
89	27	100.0	20	2	AAW16373
90	27	100.0	20	4	AAE08223
91	27	100.0	20	7	ADG18669
92	27	100.0	20	9	ADW92831
93	27	100.0	23	2	AAV01040
94	27	100.0	24	2	AAV51054
95	27	100.0	24	2	AAV74608
96	27	100.0	24	2	AAV74991
97	27	100.0	25	2	AAV41482

Aag94474	Human com
Aag94478	Human com
Aag95174	Human com
Aag95528	Human com
Aag97180	Human com
Aag94834	Human com
Aag95878	Human com
Aag96208	Human com
Adj66128	Epichloe
Aaw09224	Repeat pe
Aay01039	Methylate
Aay51894	Protocol
Aay90997	Inhibitor
Abg31424	Synthetic
Abw01640	Synthetic
Adj96780	Stabilise
Adr45935	Human hig
Adr45929	Human hig
Adr45932	Human hig
Aaw65568	Epstein-B
Aay01044	Methylate
Aae09159	Epstein-B
Abu07643	Epstein-B
Adol7059	Epstein-B
Abb82931	Arginie-d
Abb82932	Arginie-d
Abp54101	Transport
Abb82933	Arginie-d
Abb82934	Arginie-d
Abb82935	Arginie-d
Adc19846	Fluoresce
Ads52349	S33 pepti
Aap60481	Epstein-B
Aar25058	Synthetic
Abp25029	HIV DR 3b
Abp51969	Human RNA
Adf89480	Human E2F
Adf89481	Human E2F
Adk48862	C-termina
Adk48863	C-termina
Ads52340	Smd3 C-te
Ads52345	Smd3 C-te
Ads52330	Smd1 C-te
Ads52335	Smd1 C-te
Ads52343	Smd3 C-te
Ads52337	Smd1 C-te
Ads52338	Smd3 C-te
Ads52341	Smd3 C-te
Ads52334	Smd1 C-te
Ads52331	Smd1 C-te
Ads52342	Smd3 C-te
Ads52332	Smd1 C-te
Ads52339	Smd3 C-te
Ads52333	Smd1 C-te
Ads52336	Smd1 C-te
Aaw65569	Epstein-B
Aay01042	Methylate
Aae09160	Epstein-B
Aag64369	Peptide r
Abu07644	Epstein-B
Adol7080	Epstein-B
Aea01837	Membrane
Aay20639	Human neu
Aay01038	Methylate
Aaw16373	Synthetic
Aae08223	Ribonucle
Adg18669	Human scl
Adw92831	FBRL nucl
Aay01040	Methylate
Aar51054	Epstein-B
Aar74608	HIV-1 gp1
Aar74991	Epstein-B
Aay41482	Fragment

98	27	100.0	25	7	ADC99418	Adc99418	Peptide i	171	27	100.0	36	4	AAG67984	Human glu
99	27	100.0	26	2	AAG65562	Aaw65562	Epstein-B	172	27	100.0	36	4	AAO09096	Human pol
100	27	100.0	26	2	AAG65592	Aaw65592	Intgenic	173	27	100.0	37	2	AAW16625	Sm-D anti
101	27	100.0	26	4	AAE09156	Aae09156	Epstein-B	174	27	100.0	37	2	AAW63243	Arg26, 34-
102	27	100.0	26	4	AAE09156	Aae09156	Epstein-B	175	27	100.0	37	2	AAW63252	Arg26, 34-
103	27	100.0	26	6	ABU07635	Abu07635	Human aut	176	27	100.0	37	2	AAW63270	Arg26, 34-
104	27	100.0	26	6	ABU07637	Abu07637	Epstein-B	177	27	100.0	37	2	AAW63261	Arg26, 34-
105	27	100.0	26	8	ADU017080	Adol17080	Epstein-B	178	27	100.0	37	2	AAW63223	Arg26, 34-
106	27	100.0	26	8	ADU017055	Adol17055	Epstein-B	179	27	100.0	37	2	AAW63252	Arg26, 34-
107	27	100.0	28	2	AAE20163	Aar20163	Peptide #	180	27	100.0	37	2	AAW63252	GLP-1 mut
108	27	100.0	29	2	AAE20163	Aar20163	Peptide #	181	27	100.0	37	2	AAW63252	GLP-1 mut
109	27	100.0	29	4	ABBS1145	Abbs1145	Methylate	182	27	100.0	37	2	AAW63252	GLP-1 mut
110	27	100.0	29	6	ABBS1145	Abbs1145	Human sec	183	27	100.0	37	2	AAW63252	GLP-1 mut
111	27	100.0	29	6	ABO45402	Abos45402	Novel hum	184	27	100.0	37	2	AAW63252	GLP-1 mut
112	27	100.0	30	2	ABO26882	Abos26882	Protein a	185	27	100.0	37	4	ABW44192	Peptide #
113	27	100.0	31	4	ABW37781	Abw37781	Synthetic	186	27	100.0	37	4	ABW44192	Peptide #
114	27	100.0	31	4	AAW31174	Aam31174	Peptide #	187	27	100.0	37	4	AAW38145	Peptide #
115	27	100.0	31	4	ABW31174	Abw31174	Peptide #	188	27	100.0	37	4	AAW68014	Human glu
116	27	100.0	31	4	ABW31174	Abw31174	Peptide #	189	27	100.0	37	4	AAW68005	Human glu
117	27	100.0	31	4	AAW70882	Aam70882	Human bon	190	27	100.0	37	4	AAW68023	Human glu
118	27	100.0	31	4	AAW58396	Aam58396	Human bra	191	27	100.0	37	4	AAW68032	Human glu
119	27	100.0	31	4	ABG52598	Abg52598	Human liv	192	27	100.0	37	4	AAW68032	Human glu
120	27	100.0	31	5	ABG40677	Abg40677	Human pep	193	27	100.0	37	4	AAW68032	Human glu
121	27	100.0	32	2	AAE42952	Aay42952	Lys(18, 23	194	27	100.0	37	4	AAW68032	Human bon
122	27	100.0	32	2	AAE42942	Aay42942	Arg34-GLP	195	27	100.0	37	4	AAW68032	Human bra
123	27	100.0	32	2	AAE09082	Aae09082	Epstein-B	196	27	100.0	37	4	ABG59565	Human liv
124	27	100.0	33	2	AAW63213	Aaw63213	Gly8-Arg2	197	27	100.0	37	5	ABG46948	Human pep
125	27	100.0	33	2	AAW63201	Aaw63201	Arg26, 34-	198	27	100.0	37	5	ABO57617	Human gen
126	27	100.0	33	2	AAW63285	Aaw63285	Arg34-Lys	199	27	100.0	38	2	AAW16626	Sm-D anti
127	27	100.0	33	2	AAW63209	Aay34209	GLP-1 mut	200	27	100.0	38	2	AAW63262	Arg26, 34-
128	27	100.0	33	4	AAW68107	Aag68107	Human glu	201	27	100.0	38	2	AAW63271	Arg26, 34-
129	27	100.0	33	4	AAW67981	Aag67981	Human glu	202	27	100.0	38	2	AAW63224	Arg26, 34-
130	27	100.0	33	4	AAW68047	Aag68047	Human glu	203	27	100.0	38	2	AAW63235	Arg26, 34-
131	27	100.0	34	2	AAW63267	Aaw63267	Human glu	204	27	100.0	38	2	AAW63244	Arg26, 34-
132	27	100.0	34	2	AAW63214	Aaw63214	Gly8-Arg2	205	27	100.0	38	2	AAW10141	Methylate
133	27	100.0	34	2	AAW63206	Aaw63206	Arg34-Lys	206	27	100.0	38	2	AAW34280	GLP-1 mut
134	27	100.0	34	2	AAW63202	Aaw63202	Arg26, 34-	207	27	100.0	38	2	AAW34280	GLP-1 mut
135	27	100.0	34	2	AAW63218	Aaw63218	Gly8-Arg3	208	27	100.0	38	2	AAW34253	GLP-1 mut
136	27	100.0	34	2	AAW34214	Aay34214	GLP-1 mut	209	27	100.0	38	2	AAW34253	GLP-1 mut
137	27	100.0	34	2	AAW34276	Aay34276	GLP-1 mut	210	27	100.0	38	2	AAW34262	GLP-1 mut
138	27	100.0	34	2	AAW34210	Aay34210	GLP-1 mut	211	27	100.0	38	2	AAW34233	GLP-1 mut
139	27	100.0	34	2	AAW34226	Aay34226	GLP-1 mut	212	27	100.0	38	4	AAW67997	Human glu
140	27	100.0	34	2	AAW34222	Aay34222	GLP-1 mut	213	27	100.0	38	4	AAW68033	Human glu
141	27	100.0	34	4	AAW68108	Aag68108	Human glu	214	27	100.0	38	4	AAW68006	Human glu
142	27	100.0	34	4	AAW68029	Aag68029	Human glu	215	27	100.0	38	4	AAW68015	Human glu
143	27	100.0	34	4	AAW68112	Aag68112	Human glu	216	27	100.0	38	4	AAW68024	Human glu
144	27	100.0	34	4	AAW68101	Aag68101	Human glu	217	27	100.0	38	4	AAW68024	Human glu
145	27	100.0	34	4	AAW67982	Aag67982	Human glu	218	27	100.0	38	5	ADK35789	Novel hum
146	27	100.0	34	4	ABW14798	Abw14798	Human ner	219	27	100.0	39	2	AAW16627	Sm-D anti
147	27	100.0	35	2	AAW16623	Aaw16623	Sm-D anti	220	27	100.0	39	2	AAW63225	Arg26, 34-
148	27	100.0	35	2	AAW57053	Aaw57053	E2F activ	221	27	100.0	39	2	AAW63282	Arg34-Lys
149	27	100.0	35	2	AAW63259	Aaw63259	Arg26, 34-	222	27	100.0	39	2	AAW63263	Arg26, 34-
150	27	100.0	35	2	AAW63221	Aaw63221	Arg26, 34-	223	27	100.0	39	2	AAW63227	Arg26, 34-
151	27	100.0	35	2	AAW63268	Aaw63268	GLP-1 mut	224	27	100.0	39	2	AAW63236	Arg26, 34-
152	27	100.0	35	2	AAW34268	Aay34268	GLP-1 mut	225	27	100.0	39	2	AAW63254	Arg26, 34-
153	27	100.0	35	2	AAW34230	Aay34230	GLP-1 mut	226	27	100.0	39	2	AAW63272	Arg26, 34-
154	27	100.0	35	2	AAW34270	Aay34270	GLP-1 mut	227	27	100.0	39	2	AAW63245	Arg26, 34-
155	27	100.0	35	4	AAW69562	Aab69562	Human Rep	228	27	100.0	39	2	AAW63245	Arg26, 34-
156	27	100.0	35	4	AAW67983	Aag67983	Human glu	229	27	100.0	39	2	AAW34254	GLP-1 mut
157	27	100.0	35	4	AAW68021	Aag68021	Human glu	230	27	100.0	39	2	AAW34254	GLP-1 mut
158	27	100.0	35	4	AAW68030	Aag68030	Human glu	231	27	100.0	39	2	AAW34263	GLP-1 mut
159	27	100.0	36	2	AAW16624	Aaw16624	Sm-D anti	232	27	100.0	39	2	AAW34236	GLP-1 mut
160	27	100.0	36	2	AAW63260	Aaw63260	Arg26, 34-	233	27	100.0	39	2	AAW34281	GLP-1 mut
161	27	100.0	36	2	AAW63251	Aaw63251	Arg26, 34-	234	27	100.0	39	2	AAW34291	GLP-1 mut
162	27	100.0	36	2	AAW63222	Aaw63222	Arg26, 34-	235	27	100.0	39	2	AAW34291	GLP-1 mut
163	27	100.0	36	2	AAW63269	Aaw63269	Arg26, 34-	236	27	100.0	39	2	AAW34293	GLP-1 mut
164	27	100.0	36	2	AAW34260	Aay34260	GLP-1 mut	237	27	100.0	39	2	AAW34293	GLP-1 mut
165	27	100.0	36	2	AAW34278	Aay34278	GLP-1 mut	238	27	100.0	39	4	ABW32533	Peptide #
166	27	100.0	36	2	AAW34231	Aay34231	GLP-1 mut	239	27	100.0	39	4	AAW26001	Peptide #
167	27	100.0	36	2	AAW34269	Aay34269	GLP-1 mut	240	27	100.0	39	4	AAW68016	Human glu
168	27	100.0	36	4	AAW68013	Aag68013	Human glu	241	27	100.0	39	4	AAW67987	Human glu
169	27	100.0	36	4	AAW68022	Aag68022	Human glu	242	27	100.0	39	4	AAW67998	Human glu
170	27	100.0	36	4	AAW68031	Aag68031	Human glu	243	27	100.0	39	4	AAW68025	Human glu

244	27	100.0	39	4	AAG68044	Aag68044 Human glu	317	27	100.0	43	4	AAG68011	Aag68011 Human glu
245	27	100.0	39	4	AAG67989	Aag67989 Human glu	318	27	100.0	44	2	AAW16632	Aaw16632 Sm-D anti
246	27	100.0	39	4	AAG68007	Aag68007 Human glu	319	27	100.0	44	2	AAW63241	Aaw63241 Arg26, 34-
247	27	100.0	39	4	AAG68034	Aag68034 Human glu	320	27	100.0	44	2	AAW63232	Aaw63232 Arg26, 34-
248	27	100.0	39	4	ABB27385	Abb27385 Human pep	321	27	100.0	44	2	AAW34241	Aay34241 GLP-1 mut
249	27	100.0	39	4	ABB18037	Abb18037 Protein #	322	27	100.0	44	2	AAW34250	Aay34250 GLP-1 mut
250	27	100.0	39	4	AAW65745	Aam65745 Human bon	323	27	100.0	44	4	AAW19525	Aam19525 Peptide #
251	27	100.0	39	4	AAW53366	Aam53366 Human bra	324	27	100.0	44	4	ABB39044	Abb39044 Peptide #
252	27	100.0	39	4	ABG47386	Abg47386 Human liv	325	27	100.0	44	4	AAW32533	Aam32533 Peptide #
253	27	100.0	39	4	AAW01354	Aam01354 Peptide #	326	27	100.0	44	4	AAW67994	Aag67994 Human glu
254	27	100.0	39	5	ABG35374	Abg35374 Human pep	327	27	100.0	44	4	AAG68003	Aag68003 Human glu
255	27	100.0	40	2	AAW16628	Aaw16628 Sm-D anti	328	27	100.0	44	4	ABB23983	Abb23983 Protein #
256	27	100.0	40	2	AAW63255	Aaw63255 Arg26, 34-	329	27	100.0	44	4	AAW72274	Aam72274 Human bon
257	27	100.0	40	2	AAW63264	Aaw63264 Arg26, 34-	330	27	100.0	44	4	AAW59696	Aam59696 Human bra
258	27	100.0	40	2	AAW63246	Aaw63246 Arg26, 34-	331	27	100.0	44	4	ABG53960	Abg53960 Human liv
259	27	100.0	40	2	AAW63273	Aaw63273 Arg26, 34-	332	27	100.0	44	5	ABG42089	Abg42089 Human pep
260	27	100.0	40	2	AAW63228	Aaw63228 Arg26, 34-	333	27	100.0	45	2	AAW16633	Aaw16633 Sm-D anti
261	27	100.0	40	2	AAW63237	Aaw63237 Arg26, 34-	334	27	100.0	45	2	AAW63233	Aaw63233 Arg26, 34-
262	27	100.0	40	2	AAW34237	Aay34237 GLP-1 mut	335	27	100.0	45	2	AAW34242	Aay34242 GLP-1 mut
263	27	100.0	40	2	AAW34282	Aay34282 GLP-1 mut	336	27	100.0	45	4	AAW67995	Aag67995 Human glu
264	27	100.0	40	2	AAW34255	Aay34255 GLP-1 mut	337	27	100.0	46	4	ABB14818	Abb14818 Human ner
265	27	100.0	40	2	AAW34264	Aay34264 GLP-1 mut	338	27	100.0	47	6	ABP80076	Abp80076 N. Gonorr
266	27	100.0	40	2	AAW34273	Aay34273 GLP-1 mut	339	27	100.0	48	3	AAW45333	Aaw45333 Human sec
267	27	100.0	40	2	AAW34246	Aay34246 GLP-1 mut	340	27	100.0	49	4	AAW91444	Aam91444 Human imm
268	27	100.0	40	4	AAG67999	Aag67999 Human glu	341	27	100.0	49	8	ABO58117	Abo58117 Human gen
269	27	100.0	40	4	AAG68026	Aag68026 Human glu	342	27	100.0	50	5	ABP08850	Abp08850 Human ORF
270	27	100.0	40	4	AAG68008	Aag68008 Human glu	343	27	100.0	50	8	ABO53775	Abo53775 Human gen
271	27	100.0	40	4	AAG67990	Aag67990 Human glu	344	27	100.0	51	4	AAU52252	Aau52252 Proptonib
272	27	100.0	40	4	AAG68035	Aag68035 Human glu	345	27	100.0	51	5	AAE17020	Aae17020 Hepatitis
273	27	100.0	40	4	AAG68017	Aag68017 Human glu	346	27	100.0	51	6	ABM48777	Abm48777 Proptonib
274	27	100.0	41	2	AAW16629	Aaw16629 Sm-D anti	347	27	100.0	51	8	ADQ07579	Adq07579 Streptomy
275	27	100.0	41	2	AAW63247	Aaw63247 Arg26, 34-	348	27	100.0	52	2	AAW12112	Aar12112 Angigeni
276	27	100.0	41	2	AAW63256	Aaw63256 Arg26, 34-	349	27	100.0	52	3	AAW41124	Aaw41124 Zea may
277	27	100.0	41	2	AAW63265	Aaw63265 Arg26, 34-	350	27	100.0	52	4	AAU61862	Aau61862 Proptonib
278	27	100.0	41	2	AAW63229	Aaw63229 Arg26, 34-	351	27	100.0	52	6	ABW58381	Abw58381 Proptonib
279	27	100.0	41	2	AAW63238	Aaw63238 Arg26, 34-	352	27	100.0	53	2	AAW99378	Aaw99378 Human fib
280	27	100.0	41	2	AAW34285	Aay34285 GLP-1 mut	353	27	100.0	53	4	AAU49571	Aau49571 Proptonib
281	27	100.0	41	2	AAW34238	Aay34238 GLP-1 mut	354	27	100.0	53	4	ABG27737	Abg27737 Novel hum
282	27	100.0	41	2	AAW34256	Aay34256 GLP-1 mut	355	27	100.0	53	6	ABM46090	Abm46090 Proptonib
283	27	100.0	41	2	AAW34274	Aay34274 GLP-1 mut	356	27	100.0	54	2	AAW35981	Aaw35981 Extended
284	27	100.0	41	2	AAW34247	Aay34247 GLP-1 mut	357	27	100.0	54	4	AAU47889	Aau47889 Proptonib
285	27	100.0	41	4	AAG68009	Aag68009 Human glu	358	27	100.0	54	6	ABM44408	Abm44408 Proptonib
286	27	100.0	41	4	AAG67991	Aag67991 Human glu	359	27	100.0	54	8	ADP19289	Adp19289 Human sec
287	27	100.0	41	4	AAG68018	Aag68018 Human glu	360	27	100.0	55	4	ABM40725	Abm40725 Peptide #
288	27	100.0	41	4	AAG68007	Aag68007 Human glu	361	27	100.0	55	4	AAW34491	Aam34491 Peptide #
289	27	100.0	41	4	AAG68027	Aag68027 Human glu	362	27	100.0	55	4	AAW74378	Aam74378 Human bon
290	27	100.0	41	8	ABO54033	Abo54033 Human gen	363	27	100.0	55	4	AAW61587	Aam61587 Human bra
291	27	100.0	42	2	AAW16630	Aaw16630 Sm-D anti	364	27	100.0	55	4	ABG56173	Abg56173 Human liv
292	27	100.0	42	2	AAW63248	Aaw63248 Arg26, 34-	365	27	100.0	55	5	ABP03130	Abp03130 Human ORF
293	27	100.0	42	2	AAW63257	Aaw63257 Arg26, 34-	366	27	100.0	55	5	ABG44285	Abg44285 Human pep
294	27	100.0	42	2	AAW63230	Aaw63230 Arg26, 34-	367	27	100.0	55	8	ADG22457	Adg22457 Cyanophag
295	27	100.0	42	2	AAW63239	Aaw63239 Arg26, 34-	368	27	100.0	55	8	ADP45506	Adp45506 Human col
296	27	100.0	42	2	AAW00315	Aay00315 Human sec	369	27	100.0	56	2	AAW99375	Aaw99375 Human fib
297	27	100.0	42	2	AAW34248	Aay34248 GLP-1 mut	370	27	100.0	56	4	AAU49495	Aau49495 Proptonib
298	27	100.0	42	2	AAW34239	Aay34239 GLP-1 mut	371	27	100.0	56	5	ABP33945	Abp33945 Human tra
299	27	100.0	42	2	AAW34266	Aay34266 GLP-1 mut	372	27	100.0	56	6	ABM46014	Abm46014 Proptonib
300	27	100.0	42	2	AAW34257	Aay34257 GLP-1 mut	373	27	100.0	56	8	ABO58563	Abo58563 Human gen
301	27	100.0	42	4	AAG68001	Aag68001 Human glu	374	27	100.0	57	4	ABM82958	Aam82958 Human imm
302	27	100.0	42	4	AAG67992	Aag67992 Human glu	375	27	100.0	57	4	AAU48605	Aau48605 Proptonib
303	27	100.0	42	4	AAG68019	Aag68019 Human glu	376	27	100.0	57	6	ABM45124	Abm45124 Proptonib
304	27	100.0	42	4	AAG68010	Aag68010 Human glu	377	27	100.0	58	2	AAW99377	Aaw99377 Human fib
305	27	100.0	42	7	ADP90345	Adp90345 Novel hum	378	27	100.0	58	4	AAU49149	Aau49149 Proptonib
306	27	100.0	42	7	ADG90164	Adg90164 Human pro	379	27	100.0	58	4	AAU46501	Aau46501 Proptonib
307	27	100.0	42	9	ADY25504	Ady25504 Novel hum	380	27	100.0	58	6	ABM45668	Abm45668 Human imm
308	27	100.0	43	2	AAW16631	Aaw16631 Sm-D anti	381	27	100.0	58	6	ABM43020	Abm43020 Proptonib
309	27	100.0	43	2	AAW63231	Aaw63231 Arg26, 34-	382	27	100.0	59	4	AAO50987	Aao50987 Epstein-B
310	27	100.0	43	2	AAW63240	Aaw63240 Arg26, 34-	383	27	100.0	60	4	AAW17151	Aam17151 Peptide #
311	27	100.0	43	2	AAW63249	Aaw63249 Arg26, 34-	384	27	100.0	60	4	ABB36152	Abb36152 Peptide #
312	27	100.0	43	2	AAW34240	Aay34240 GLP-1 mut	385	27	100.0	60	4	AAW29643	Aam29643 Peptide #
313	27	100.0	43	2	AAW34258	Aay34258 GLP-1 mut	386	27	100.0	60	4	ABB30961	Abb30961 Peptide #
314	27	100.0	43	2	AAW34249	Aay34249 GLP-1 mut	387	27	100.0	60	4	ABB21539	Abb21539 Protein #
315	27	100.0	43	4	AAG68002	Aag68002 Human glu	388	27	100.0	60	4	AAU62807	Aau62807 Proptonib
316	27	100.0	43	4	AAG67993	Aag67993 Human glu	389	27	100.0	60	4	AAW56937	Aam56937 Human bra

390	27	100.0	60	4	ABG50997	Abg50997 Human liv	463	27	100.0	75	6	ABM46528	Abm46528 Propionib
391	27	100.0	60	4	AAm04846	Aam04846 Peptide #	464	27	100.0	76	4	ABG06899	Abg06899 Novel hum
392	27	100.0	60	4	AAb76874	Aab76874 Human lun	465	27	100.0	77	4	AAm78240	Aam78240 Human bon
393	27	100.0	60	5	AAu85529	Aau85529 L801P lun	466	27	100.0	77	4	ABE14769	Abel14769 Human ner
394	27	100.0	60	6	ABM59326	Abm59326 Propionib	467	27	100.0	77	5	ABG47253	Abg47253 Human pep
395	27	100.0	60	6	ABU69501	Abu69501 Human lun	468	27	100.0	77	8	ADP46660	Adp46660 Human col
396	27	100.0	60	6	ABU66403	Abu66403 Lung canc	469	27	100.0	78	4	AAU39381	Aau39381 Propionib
397	27	100.0	60	7	ADH46318	Adh46318 Human lun	470	27	100.0	78	6	ABM35900	Abm35900 Propionib
398	27	100.0	60	8	ADJ20237	Adj20237 Human lun	471	27	100.0	79	3	AAg18775	Aag18775 Zea maye
399	27	100.0	61	8	ABO58258	Abos8258 Human gen	472	27	100.0	80	3	AAg26025	Aag26025 Zea maye
400	27	100.0	62	4	ABG16483	Abg16483 Novel hum	473	27	100.0	81	3	AAg11733	Aag11733 Arabidops
401	27	100.0	63	3	AAb44937	Aab44937 Human sec	474	27	100.0	81	3	AAg22637	Aag22637 Zea maye
402	27	100.0	63	4	AAE60725	Aae60725 Human sec	475	27	100.0	81	4	AAU44950	Aau44950 Propionib
403	27	100.0	63	8	ABO55542	Abos5542 Human gen	476	27	100.0	81	6	ABM41469	Abm41469 Propionib
404	27	100.0	64	4	AAU27874	Aau27874 Human con	477	27	100.0	83	3	ABG406027	Abg406027 Arabidops
405	27	100.0	64	4	AAU56453	Aau56453 Propionib	478	27	100.0	83	8	ADR45919	Adr45919 Human hig
406	27	100.0	64	4	ABE14824	Abel14824 Human ner	479	27	100.0	84	3	AAg60047	Aag60047 Arabidops
407	27	100.0	64	6	ABM52972	Abm52972 Propionib	480	27	100.0	85	4	AAm85702	Aam85702 Human inm
408	27	100.0	66	8	ABO53906	Abos3906 Human gen	481	27	100.0	85	4	AAU53552	Aau53552 Propionib
409	27	100.0	67	2	AAW99376	Aaw99376 Human fib	482	27	100.0	85	4	ABG23437	Abg23437 Novel hum
410	27	100.0	67	4	AAU46954	Aau46954 Propionib	483	27	100.0	85	5	AAE23312	Aae23312 Human p54
411	27	100.0	67	5	ABP07125	Abp07125 Human ORF	484	27	100.0	85	5	AAE23318	Aae23318 Human p85
412	27	100.0	67	6	ABM43473	Abm43473 Propionib	485	27	100.0	85	5	ADJ33795	Adj33795 Proteín o
413	27	100.0	68	5	ABJ03708	Abj03708 Human ova	486	27	100.0	85	6	ABM50071	Abm50071 Propionib
414	27	100.0	68	8	ADX94889	Adx94889 Plant ful	487	27	100.0	86	4	AAg90602	Aag90602 C glutami
415	27	100.0	69	4	ABB14787	Abb14787 Human ner	488	27	100.0	86	4	ABG02078	Abg02078 Novel hum
416	27	100.0	70	4	ABG09104	Abg09104 Novel hum	489	27	100.0	86	7	ADJ85800	Adj85800 Truncated
417	27	100.0	70	8	ABO57495	Abos7495 Human gen	490	27	100.0	87	4	AAm24476	Aam24476 Human EST
418	27	100.0	71	4	AAm21312	Aam21312 Peptide #	491	27	100.0	87	4	ABB16767	Abb16767 Human ner
419	27	100.0	71	4	ABB43645	Abb43645 Peptide #	492	27	100.0	87	8	ADP81176	Adp81176 Protein o
420	27	100.0	71	4	ABB40920	Abb40920 Peptide #	493	27	100.0	87	8	ADY13468	Ady13468 Plant ful
421	27	100.0	71	4	AAm34692	Aam34692 Peptide #	494	27	100.0	88	7	ADE85801	Ade85801 Truncated
422	27	100.0	71	4	AAm74580	Aam74580 Human bon	495	27	100.0	89	2	AAW30064	Aaw30064 Mature mo
423	27	100.0	71	4	AAU41429	Aau41429 Propionib	496	27	100.0	89	2	AAy16661	Aay16661 Murine pe
424	27	100.0	71	4	AAU64700	Aau64700 Propionib	497	27	100.0	89	4	AAU66074	Aau66074 Propionib
425	27	100.0	71	4	AAm61782	Aam61782 Human bra	498	27	100.0	89	4	AAU58281	Aau58281 Propionib
426	27	100.0	71	4	AAm64589	Aam64589 Human bra	499	27	100.0	89	4	AAU51733	Aau51733 Propionib
427	27	100.0	71	4	ABG56367	Abg56367 Human liv	500	27	100.0	89	4	AAU03925	Aau03925 Murine pa
428	27	100.0	71	4	ABG59016	Abg59016 Human liv	501	27	100.0	89	4	AAU03939	Aau03939 Murine su
429	27	100.0	71	5	ABG44419	Abg44419 Human pep	502	27	100.0	89	6	ABM48252	Abm48252 Propionib
430	27	100.0	71	5	ABG46402	Abg46402 Human pep	503	27	100.0	89	6	ABM54800	Abm54800 Propionib
431	27	100.0	71	6	ABM37948	Abm37948 Propionib	504	27	100.0	89	6	ABM62593	Abm62593 Propionib
432	27	100.0	71	6	AAm61219	Aam61219 Propionib	505	27	100.0	90	4	AAU47727	Aau47727 Propionib
433	27	100.0	72	4	AAm16479	Aam16479 Peptide #	506	27	100.0	90	4	AAU03935	Aau03935 Murine PG
434	27	100.0	72	4	ABB35464	Abb35464 Peptide #	507	27	100.0	90	4	AAU03937	Aau03937 Murine PN
435	27	100.0	72	4	AAm28969	Aam28969 Peptide #	508	27	100.0	90	4	AAU32687	Aau32687 Novel hum
436	27	100.0	72	4	ABB30297	Abb30297 Peptide #	509	27	100.0	90	6	ABM44246	Abm44246 Propionib
437	27	100.0	72	4	ABE20901	Abel20901 Protein #	510	27	100.0	90	8	ADR45920	Adr45920 Human hig
438	27	100.0	72	4	AAm68662	Aam68662 Human bon	511	27	100.0	91	2	AAy16668	Aay16668 WO9914235
439	27	100.0	72	4	AAU48582	Aau48582 Human pol	512	27	100.0	91	3	AAg34737	Aag34737 Arabidops
440	27	100.0	72	4	AAO09150	Aao09150 Human pol	513	27	100.0	91	4	AAm86465	Aam86465 Human inm
441	27	100.0	72	4	AAU65970	Aau65970 Propionib	514	27	100.0	91	4	ABG26453	Abg26453 Novel hum
442	27	100.0	72	4	AAU44238	Aau44238 Propionib	515	27	100.0	91	4	ABG14107	Abg14107 Novel hum
443	27	100.0	72	4	AAm56285	Aam56285 Human bra	516	27	100.0	91	4	ABG17761	Abg17761 Novel hum
444	27	100.0	72	4	AAU50323	Aau50323 Human liv	517	27	100.0	91	4	ABG17734	Abg17734 Novel hum
445	27	100.0	72	4	AAm04206	Aam04206 Peptide #	518	27	100.0	92	4	AAO13040	Aao13040 Human pol
446	27	100.0	72	5	ABP09144	Abp09144 Human ORF	519	27	100.0	92	8	ADR45925	Adr45925 Human hig
447	27	100.0	72	5	ABG38242	Abg38242 Human pep	520	27	100.0	93	3	AAg34735	Aag34735 Arabidops
448	27	100.0	72	6	ABM45101	Abm45101 Propionib	521	27	100.0	93	4	AAU59369	Aau59369 Propionib
449	27	100.0	72	6	ABM62289	Abm62289 Propionib	522	27	100.0	93	4	ABG26509	Abg26509 Novel hum
450	27	100.0	72	6	ABM40757	Abm40757 Propionib	523	27	100.0	93	6	ABM64814	Abm64814 Propionib
451	27	100.0	72	7	ADJ09092	Adj09092 Novel pro	524	27	100.0	93	6	ABM55888	Abm55888 Propionib
452	27	100.0	73	8	ABO60213	Abos60213 Human gen	525	27	100.0	94	3	AAO6017	Aao6017 Arabidops
453	27	100.0	73	8	ABO56446	Abos6446 Human gen	526	27	100.0	94	4	AAO08238	Aao08238 Human pol
454	27	100.0	74	3	AAy68733	Aay68733 Amino aci	527	27	100.0	94	4	AAU41647	Aau41647 Propionib
455	27	100.0	74	4	AAU40172	Aau40172 Propionib	528	27	100.0	94	4	ABG26494	Abg26494 Novel hum
456	27	100.0	74	5	ADP04336	Adp04336 Human ORF	529	27	100.0	94	6	ABM38166	Abm38166 Propionib
457	27	100.0	74	5	ADK34381	Adk34381 Novel hum	530	27	100.0	95	5	ABP64293	Abp64293 Human ORF
458	27	100.0	74	5	ADK34380	Adk34380 Novel hum	531	27	100.0	95	8	ADR45928	Adr45928 Human hig
459	27	100.0	74	6	ABM36691	Abm36691 Propionib	532	27	100.0	95	9	ABM94063	Abm94063 M. xanthu
460	27	100.0	75	4	AAU50009	Aau50009 Propionib	533	27	100.0	96	2	AAW30075	Aaw30075 Fersephin
461	27	100.0	75	4	ABG11271	Abg11271 Novel hum	534	27	100.0	96	2	AAW30066	Aaw30066 Mature mo
462	27	100.0	75	5	ABP04353	Abp04353 Human ORF	535	27	100.0	96	2	AAy16662	Aay16662 WO9914235



536	27	100.0	96	2	AA116695	Ray16695 WO9914235	609	27	100.0	107	5	AA014033	Aao14033 Human hig
537	27	100.0	96	2	AA116723	Ray16723 Mouse mat	610	27	100.0	107	5	AA014034	Aao14034 Chinese h
538	27	100.0	96	3	AA168730	Ray68730 Amino aci	611	27	100.0	107	5	ADQ97197	Adq97197 Human can
539	27	100.0	96	4	AA050718	Abg50718 Human HMG	612	27	100.0	107	8	ADQ97199	Adq97199 Human can
540	27	100.0	96	4	ABG26490	Abg26490 Novel hum	613	27	100.0	107	8	ADQ97201	Adq97201 Human can
541	27	100.0	96	4	ABG17098	Abg17098 Novel hum	614	27	100.0	107	8	ADR45916	Adr45916 Human hig
542	27	100.0	96	4	AAU03955	Aau03955 Mouse per	615	27	100.0	107	8	ADP23528	Adp23528 PRO polyp
543	27	100.0	96	5	AA014035	Aao14035 Chinese h	616	27	100.0	107	9	ADX07512	Adx07512 Cyclin-de
544	27	100.0	96	5	AA014036	Aao14036 Mouse hig	617	27	100.0	107	9	ADZ09817	Adz09817 Human bre
545	27	100.0	96	8	ADQ97195	Adq97195 Human can	618	27	100.0	108	2	AAW64658	Aaw64658 Chicken H
546	27	100.0	96	8	ADR45917	Adr45917 Human hig	619	27	100.0	108	3	AAG01102	Aag01102 Human sec
547	27	100.0	96	8	ADR45926	Adr45926 Human hig	620	27	100.0	108	4	AAU64015	Aau64015 Propionib
548	27	100.0	96	8	ADR45921	Adr45921 Human hig	621	27	100.0	108	4	AAU03938	Aau03938 Murine Hi
549	27	100.0	96	8	ADP23951	Adp23951 PRO polyp	622	27	100.0	108	5	ABF08307	Abf08307 Human ORF
550	27	100.0	96	9	ADW72114	Adw72114 Human HYG	623	27	100.0	108	6	ABM60534	Abm60534 Propionib
551	27	100.0	96	9	AAU07510	Adu07510 Cyclin-de	624	27	100.0	108	7	ADG36879	Adg36879 Mouse hig
552	27	100.0	97	4	AAU67945	Aau67945 Propionib	625	27	100.0	108	2	AAW02134	Aaw02134 High mobi
553	27	100.0	97	6	ABM64464	Abm64464 Propionib	626	27	100.0	109	3	AAG11732	Aag11732 Arabidops
554	27	100.0	97	8	ADR95182	Adr95182 Novel S.	627	27	100.0	109	3	AAG22636	Aag22636 Zea may
555	27	100.0	97	8	ADY04747	Ady04747 Plant ful	628	27	100.0	109	3	AAG40841	Aag40841 Zea may
556	27	100.0	97	8	ADY23616	Ady23616 Plant ful	629	27	100.0	109	4	ABG90767	Abg90767 Human she
557	27	100.0	97	9	AEA59052	Aea59052 Streptoco	630	27	100.0	109	4	ABG26507	Abg26507 Novel hum
558	27	100.0	98	3	AAG27080	Aag27080 Zea may	631	27	100.0	109	4	AAU03934	Aau03934 Murine Hi
559	27	100.0	98	8	ADY11770	Ady11770 plant ful	632	27	100.0	109	4	AAU03936	Aau03936 Murine Hi
560	27	100.0	98	9	ABM93370	Abm93370 M. xanthu	633	27	100.0	109	7	ADG36880	Adg36880 Human hig
561	27	100.0	99	4	AAU18386	Aau18386 Peptide #	634	27	100.0	109	7	ADN95675	Adn95675 Human BEC
562	27	100.0	99	4	ABB37417	Abb37417 Peptide #	635	27	100.0	109	8	ADR45918	Adr45918 Human hig
563	27	100.0	99	4	AAU030865	Aau030865 Peptide #	636	27	100.0	110	4	ABB11689	Abb11689 Human IKK
564	27	100.0	99	4	ABB32171	Abb32171 Peptide #	637	27	100.0	110	4	AAU04453	Aau04453 Murine mu
565	27	100.0	99	4	ABB22715	Abb22715 Protein #	638	27	100.0	111	3	AAG36028	Aag36028 Zea may
566	27	100.0	99	4	AAU70545	Aau70545 Human bon	639	27	100.0	111	4	AAU23544	Aau23544 Novel hum
567	27	100.0	99	4	AAU40547	Aau40547 Propionib	640	27	100.0	111	4	ADG27735	Adg27735 Human nov
568	27	100.0	99	4	AAU58098	Aau58098 Human bra	641	27	100.0	111	8	ADT58858	Adt58858 Plant pol
569	27	100.0	99	4	ABG52224	Abg52224 Human liv	642	27	100.0	111	8	ADY12721	Ady12721 Plant ful
570	27	100.0	99	4	AAU05983	Aau05983 Peptide #	643	27	100.0	112	3	AAG03937	Aag03937 Human sec
571	27	100.0	99	5	ABG40200	Abg40200 Human pep	644	27	100.0	112	3	AAG17536	Aag17536 Arabidops
572	27	100.0	99	6	ABM37066	Abm37066 Propionib	645	27	100.0	112	4	ABE51144	Abbe51144 Human sec
573	27	100.0	99	9	ADY30588	Ady30588 Human spl	646	27	100.0	112	4	ABG03677	Abg03677 Human mus
574	27	100.0	100	2	AAU08501	Aau08501 Human tan	647	27	100.0	112	4	ABG03606	Abg03606 Novel hum
575	27	100.0	100	2	AAU05634	Aau05634 HIV-1 gro	648	27	100.0	112	6	ABU12971	Abu12971 Novel hum
576	27	100.0	100	2	AAU05648	Aau05648 HIV-1 gro	649	27	100.0	112	6	ABO45401	Abou45401 Novel hum
577	27	100.0	100	2	AAU05641	Aau05641 HIV-1 gro	650	27	100.0	112	7	ABO26881	Abou26881 Protein a
578	27	100.0	100	3	AAG54632	Aag54632 Zea may	651	27	100.0	112	8	ADJ28997	Adj28997 Human mus
579	27	100.0	100	4	AAU85254	Aau85254 Human imm	652	27	100.0	114	3	AAG18773	Aag18773 Zea may
580	27	100.0	100	4	AAU13259	Aau13259 Human pol	653	27	100.0	114	3	AAG54593	Aag54593 Zea may
581	27	100.0	100	6	ADA54553	Ada54553 Human pro	654	27	100.0	114	3	AAG54593	Aag54593 Arabidops
582	27	100.0	100	8	ADP83395	Adp83395 Breast ep	655	27	100.0	114	4	AAG06886	Aag06886 Arabidops
583	27	100.0	101	4	ABBI4788	Abbi4788 Human ner	656	27	100.0	114	4	ABG06096	Abg06096 Human bas
584	27	100.0	101	4	ABG29446	Abg29446 Novel hum	657	27	100.0	114	8	ADY05700	Ady05700 Plant ful
585	27	100.0	102	4	AAU89829	Aau89829 Human imm	658	27	100.0	115	3	AAG04598	Aag04598 Arabidops
586	27	100.0	102	4	AAU14728	Aau14728 Novel bon	659	27	100.0	115	3	AAG06016	Aag06016 Arabidops
587	27	100.0	102	7	ADC32952	Adc32952 Human nov	660	27	100.0	115	7	ABO72514	Abou72514 Pseudomon
588	27	100.0	102	8	ADM87797	Adm87797 Human EST	661	27	100.0	115	8	ADT57446	Adt57446 Plant pol
589	27	100.0	103	3	AAU58914	Aau58914 Breast an	662	27	100.0	115	8	ADX71788	Adx71788 Plant ful
590	27	100.0	104	4	AAU30579	Aau30579 Novel hum	663	27	100.0	116	3	AAG10840	Aag10840 Zea may
591	27	100.0	104	8	ADQ97191	Adq97191 Human can	664	27	100.0	116	3	AAG11731	Aag11731 Arabidops
592	27	100.0	104	8	ADX73765	Adx73765 Plant ful	665	27	100.0	116	8	ADT58857	Adt58857 Plant pol
593	27	100.0	104	8	ADX71589	Adx71589 Plant ful	666	27	100.0	117	6	ABU20230	Abu20230 Protein e
594	27	100.0	105	3	AAU28189	Aau28189 Human tra	667	27	100.0	117	7	ADT872470	Adt872470 Human end
595	27	100.0	105	3	AAG27079	Aag27079 Zea may	668	27	100.0	118	3	AAG21595	Aag21595 Arabidops
596	27	100.0	105	3	AAG54668	Aag54668 Arabidops	669	27	100.0	118	3	AAG15771	Aag15771 Arabidops
597	27	100.0	105	3	AAG26609	Aag26609 Arabidops	670	27	100.0	118	3	AAG25202	Aag25202 Arabidops
598	27	100.0	105	7	ABO63835	Abou63835 Klebsiell	671	27	100.0	118	8	ADR45927	Adr45927 Human hig
599	27	100.0	105	8	ABO58643	Abou58643 Human gen	672	27	100.0	118	8	ADG68961	Adg68961 Plant ful
600	27	100.0	105	8	ABO58643	Abou58643 Human gen	673	27	100.0	119	4	ABG32232	Abg32232 Novel hum
601	27	100.0	106	2	AAU26537	Aau26537 Trypanoso	674	27	100.0	119	5	ABP65255	Abp65255 Hypoxia-z
602	27	100.0	106	2	AAU23299	Aau23299 Trypanoso	675	27	100.0	119	5	ADK34444	Adk34444 Novel hum
603	27	100.0	106	7	ADF13915	Adf13915 Human end	676	27	100.0	119	6	ADA83869	Ada83869 Human POM
604	27	100.0	106	8	ADR45924	Adr45924 Human hig	677	27	100.0	119	6	ABO52913	Abou52913 Human spl
605	27	100.0	107	3	AAU68729	Aau68729 Amino aci	678	27	100.0	119	8	ADQ57325	Adq57325 Kidney de
606	27	100.0	107	3	AAG18774	Aag18774 Zea may	679	27	100.0	119	8	ADQ19307	Adq19307 Human PRO
607	27	100.0	107	3	AAG06887	Aag06887 Arabidops	680	27	100.0	119	8	ADP54370	Adp54370 Human PRO
608	27	100.0	107	4	AAO00581	Aao00581 Human pol	681	27	100.0	119	9	ADY14872	Ady14872 PRO polyp

682	27	100.0	119	9	AD270424	Adz70424	Human pro	755	27	100.0	139	8	ADP29728	Adp29728	Human sec
683	27	100.0	120	6	ABJ26381	Abj26381	Aspergill	756	27	100.0	140	3	AG21593	Ag21593	Arabidops
684	27	100.0	120	6	ABJ25781	Abj25781	Aspergill	757	27	100.0	140	3	AG15769	Ag15769	Arabidops
685	27	100.0	121	3	AG54592	Ag54592	Zea mays	758	27	100.0	140	4	ABG26498	Abg26498	Novel hum
686	27	100.0	122	4	ABG26311	Abg26311	Novel hum	759	27	100.0	140	8	ADT57453	Adt57453	Plant pol
687	27	100.0	123	2	AAR51053	Aar51053	Epstein-B	760	27	100.0	142	2	AA116681	Aa116681	WO9914235
688	27	100.0	123	2	AAR74990	Aar74990	Epstein-B	761	27	100.0	142	3	AA54761	Aa54761	Arabidops
689	27	100.0	123	4	AAU62227	Aau62227	Propionib	762	27	100.0	142	4	ABM11594	Abm11594	Human len
690	27	100.0	123	6	ABM58746	Abm58746	Propionib	763	27	100.0	142	4	AA95718	Aa95718	Human pro
691	27	100.0	123	8	ADY10264	Ady10264	Plant ful	764	27	100.0	142	7	ADC31805	Adc31805	Human nov
692	27	100.0	124	3	AA26608	Aa26608	Arabidops	765	27	100.0	142	7	ADD01167	Add01167	Human nuc
693	27	100.0	124	4	ABB72046	Abb72046	Drosophil	766	27	100.0	142	8	ABM80237	Abm80237	Tumour-as
694	27	100.0	124	4	ABG26488	Abg26488	Novel hum	767	27	100.0	142	8	ADX71013	Adx71013	Plant ful
695	27	100.0	124	6	ABU48496	Abu48496	Protein e	768	27	100.0	143	3	AA64678	Aa64678	Human 5'
696	27	100.0	125	8	ADY04992	Ady04992	Plant ful	769	27	100.0	143	4	AA50715	Aa50715	Human HMG
697	27	100.0	125	9	ADY66468	Ady66468	S. mansoni	770	27	100.0	143	7	ABO66781	AbO66781	Klebsiell
698	27	100.0	126	3	AA300873	Aa300873	Human sec	771	27	100.0	143	8	ADU72242	Adu72242	Signal pe
699	27	100.0	126	6	ABO52915	AbO52915	Human spl	772	27	100.0	143	9	AD273233	Ad273233	Human inc
700	27	100.0	126	7	AD447394	Ad447394	Rat Prote	773	27	100.0	145	5	ABR40543	AbR40543	Human sec
701	27	100.0	126	7	ADJ68816	Adj68816	Human hea	774	27	100.0	145	5	ABR40464	AbR40464	Human sec
702	27	100.0	126	7	ABO75510	AbO75510	Pseudomon	775	27	100.0	145	5	ABO83749	AbO83749	Pseudomon
703	27	100.0	126	8	ABN03878	Abn03878	Antipsozi	776	27	100.0	145	7	ABM89478	Abm89478	Rice abio
704	27	100.0	126	8	ABM82328	Abm82328	Tumour-as	777	27	100.0	146	3	AA36026	Aa36026	Zea mays
705	27	100.0	126	8	ADP54356	Adp54356	Human PRO	778	27	100.0	146	4	AAU67718	Aau67718	Propionib
706	27	100.0	126	9	ADY21341	Ady21341	Human ant	779	27	100.0	146	6	ABM64237	Abm64237	Propionib
707	27	100.0	126	9	ADY14828	Ady14828	PRO polyp	780	27	100.0	147	4	ABM69032	Abm69032	Drosophil
708	27	100.0	127	4	AAU30580	Aau30580	Novel hum	781	27	100.0	147	8	AD45923	Ad45923	Human hig
709	27	100.0	127	7	AD509014	Ad509014	Novel pro	782	27	100.0	148	4	AAU50579	Aau50579	Propionib
710	27	100.0	128	3	AA333047	Aa333047	Pinus rad	783	27	100.0	148	6	ABM47098	Abm47098	Propionib
711	27	100.0	128	4	AAU64965	Aau64965	Propionib	784	27	100.0	148	8	ADQ37053	Adq37053	Cell prol
712	27	100.0	128	6	ABM65061	Abm65061	Propionib	785	27	100.0	149	3	AB51680	Ab51680	Human sec
713	27	100.0	128	6	ABM61484	Abm61484	Propionib	786	27	100.0	149	5	ABP51455	AbP51455	Human MDD
714	27	100.0	128	8	ADY13541	Ady13541	Plant ful	787	27	100.0	149	7	ABO75836	AbO75836	Pseudomon
715	27	100.0	128	9	AA30589	Aa30589	Human epl	788	27	100.0	149	8	ADX73990	Adx73990	Plant ful
716	27	100.0	129	3	AA306015	Aa306015	Arabidops	789	27	100.0	150	8	ADY23391	Ady23391	Plant ful
717	27	100.0	129	4	ABB60099	Abb60099	Drosophil	790	27	100.0	150	8	ADY05192	AdY05192	Plant ful
718	27	100.0	129	4	ABG17100	Abg17100	Novel hum	791	27	100.0	151	2	AAU00949	Aau00949	Wheat Dra
719	27	100.0	129	7	ADM05330	Adm05330	Human pro	792	27	100.0	151	3	AB42220	Ab42220	Human ORP
720	27	100.0	130	3	AG27078	Ag27078	Zea mays	793	27	100.0	151	4	ABM63848	Abm63848	Drosophil
721	27	100.0	130	5	ADK36212	Adk36212	Novel hum	794	27	100.0	151	4	ABG26319	Abg26319	Novel hum
722	27	100.0	130	8	ADY77736	Ady77736	Plant pol	795	27	100.0	151	7	ADF58917	Adf58917	Human pol
723	27	100.0	130	8	ADY12535	Ady12535	Plant ful	796	27	100.0	151	8	ADR08870	Adr08870	Human pro
724	27	100.0	131	3	AG15770	Ag15770	Arabidops	797	27	100.0	153	3	AG08125	Ag08125	Arabidops
725	27	100.0	131	3	AG221594	Ag221594	Arabidops	798	27	100.0	154	3	AA551679	Aa551679	Gene 14.
726	27	100.0	132	2	AAU08500	Aau08500	Human tan	799	27	100.0	154	6	ABG99988	AbG99988	Human nov
727	27	100.0	132	3	AG36027	Ag36027	Zea mays	800	27	100.0	154	8	ADY04948	AdY04948	Human MDD
728	27	100.0	132	8	ADX95542	Adx95542	Plant ful	801	27	100.0	154	8	ADY04948	AdY04948	Plant ful
729	27	100.0	133	4	ABG16144	Abg16144	Novel hum	802	27	100.0	155	3	AA15602	Aa15602	Arabidops
730	27	100.0	134	2	AAW30067	Aaw30067	Mouse per	803	27	100.0	155	4	AB63729	Ab63729	Human pro
731	27	100.0	134	2	AA116663	Aa116663	WO9914235	804	27	100.0	155	4	AAO10339	Aao10339	Human pol
732	27	100.0	134	3	AAU75995	Aau75995	Human ski	805	27	100.0	155	4	AAU65423	Aau65423	Propionib
733	27	100.0	134	3	AAU75995	Aau75995	Arabidops	806	27	100.0	155	6	ABM61942	Abm61942	Propionib
734	27	100.0	134	4	AAU55934	Aau55934	Skin cell	807	27	100.0	155	8	ADG22471	Adg22471	Cyanophag
735	27	100.0	134	4	AAU63313	Aau63313	Propionib	808	27	100.0	155	8	ADQ19025	Adq19025	Human sof
736	27	100.0	134	5	ABU72134	Abu72134	Human pro	809	27	100.0	155	8	ADX93047	Adx93047	Plant ful
737	27	100.0	134	5	ABM59832	Abm59832	Propionib	810	27	100.0	155	8	ADX66647	Adx66647	Plant ful
738	27	100.0	134	6	ABR41724	AbR41724	Human DIT	811	27	100.0	156	2	AA116721	Aa116721	Murine pr
739	27	100.0	134	7	ABO74357	AbO74357	Pseudomon	812	27	100.0	156	3	AG08124	Ag08124	Arabidops
740	27	100.0	134	8	ADT60118	Adt60118	Plant pol	813	27	100.0	156	5	ABP69309	Abp69309	Human pol
741	27	100.0	135	4	AB93957	Ab93957	Human pro	814	27	100.0	156	8	ADX65760	Adx65760	Plant ful
742	27	100.0	135	4	ABG02058	Abg02058	Novel hum	815	27	100.0	157	6	ABO00570	AbO00570	Novel hum
743	27	100.0	135	7	ADU69719	Adu69719	Human hea	816	27	100.0	157	7	ABO79090	AbO79090	Pseudomon
744	27	100.0	136	4	ABG16397	Abg16397	Novel hum	817	27	100.0	157	7	ABM89898	Abm89898	Rice abio
745	27	100.0	136	4	ABG16397	Abg16397	Novel hum	818	27	100.0	158	2	AA121432	Aa121432	Human hig
746	27	100.0	136	7	ABM89662	Abm89662	Rice abio	819	27	100.0	158	2	AAU08499	Aau08499	Human tan
747	27	100.0	137	4	ABG19933	Abg19933	Novel hum	820	27	100.0	158	4	ABG09338	Abg09338	Novel hum
748	27	100.0	137	7	AD33285	Ad33285	Human nov	821	27	100.0	158	4	ABG04427	AbG04427	Novel hum
749	27	100.0	137	7	ADD27401	Add27401	Human adi	822	27	100.0	158	4	ABG27645	AbG27645	Novel hum
750	27	100.0	137	7	ADY75714	Ady75714	Pseudomon	823	27	100.0	158	8	ADY07621	AdY07621	Plant ful
751	27	100.0	137	8	ADX90208	Adx90208	Plant ful	824	27	100.0	158	9	ADX06393	AdX06393	Cyclin-de
752	27	100.0	138	7	ABO71385	AbO71385	Pseudomon	825	27	100.0	159	2	AAU08070	Aau08070	Human tan
753	27	100.0	139	3	AA22635	Aa22635	Zea mays	826	27	100.0	159	7	ADC88481	Adc88481	Ribosomal
754	27	100.0	139	7	ABO70675	AbO70675	Pseudomon	827	27	100.0	159	7	ABO83166	AbO83166	Pseudomon

828	27	100.0	159	8	ABO54101	Abos4101 Human gen	901	27	100.0	177	8	ADO28724	Ado28724 Human alp
829	27	100.0	159	8	ADT58670	Adt58670 Plant pol	902	27	100.0	177	8	ADO5228	Ado5228 Adrenerg
830	27	100.0	159	8	ADT71131	Adt71131 Plant ful	903	27	100.0	177	8	ADJ21116	Adj21116 Lung canc
831	27	100.0	160	3	AGA40839	Aga40839 Zea mays	904	27	100.0	177	9	ADX44543	Adx44543 Human cla
832	27	100.0	160	4	ABG22324	Abg22324 Novel hum	905	27	100.0	177	9	ADY83779	Ady83779 Alpha-1B-
833	27	100.0	161	3	AGA44856	Aga44856 Zea mays	906	27	100.0	177	9	AEA13816	Aea13816 Human alp
834	27	100.0	161	4	ABJ32841	Abj32841 Novel hum	907	27	100.0	178	7	ABO71569	AbO71569 Pseudomon
835	27	100.0	161	5	ABB77790	Abb77790 Amino aci	908	27	100.0	178	7	ABO70322	AbO70322 Pseudomon
836	27	100.0	161	8	ABO55412	AbO55412 Human gen	909	27	100.0	179	3	AAI68731	Aai68731 Amino aci
837	27	100.0	161	8	ADQ66460	Adq66460 Novel hum	910	27	100.0	179	4	AAI18625	Aai18625 Peptide #
838	27	100.0	162	7	ABO76129	AbO76129 Pseudomon	911	27	100.0	179	4	ABB37682	Abb37682 Peptide #
839	27	100.0	163	7	ADF13917	Adf13917 Human end	912	27	100.0	179	4	AAI31085	Aai31085 Peptide #
840	27	100.0	163	7	ABO76331	AbO76331 Pseudomon	913	27	100.0	179	4	AAI70787	Aai70787 Human bon
841	27	100.0	163	8	ADX79586	Adx79586 Plant ful	914	27	100.0	179	4	AAI58313	Aai58313 Human bra
842	27	100.0	164	3	AGA54591	Aga54591 Zea mays	915	27	100.0	179	4	ABG52497	Abg52497 Human liv
843	27	100.0	164	3	AGA32575	Abg32575 Arabidops	916	27	100.0	179	5	ABG40524	Abg40524 Human pep
844	27	100.0	164	4	AAI76873	Aab76873 Human lun	917	27	100.0	179	8	ABM81475	Abm81475 Tumour-as
845	27	100.0	164	5	AAU85528	Aau85528 L801P lun	918	27	100.0	181	7	ABO77736	AbO77736 Pseudomon
846	27	100.0	164	6	ABU69500	Abu69500 Human lun	919	27	100.0	181	8	ADX91410	Adx91410 Plant ful
847	27	100.0	164	6	ABU66402	Abu66402 Lung canc	920	27	100.0	181	8	ADY07270	Ady07270 Plant ful
848	27	100.0	164	7	ADH46317	Adh46317 Human lun	921	27	100.0	181	8	ADX95687	Adx95687 Plant ful
849	27	100.0	164	7	ABO73146	AbO73146 Pseudomon	922	27	100.0	181	8	ADY05750	Ady05750 Plant ful
850	27	100.0	164	7	ABO71347	AbO71347 Pseudomon	923	27	100.0	182	7	ABO61784	AbO61784 Klebsiell
851	27	100.0	164	8	ADJ20236	Adj20236 Human lun	924	27	100.0	183	8	ADY11788	Ady11788 Plant ful
852	27	100.0	164	8	ADY04503	Ady04503 Plant ful	925	27	100.0	183	8	ADY10114	Ady10114 Plant ful
853	27	100.0	165	7	ABO70663	AbO70663 Pseudomon	926	27	100.0	184	4	AAI63252	Aai63252 Human bre
854	27	100.0	166	2	AAI08498	Aay08498 Human tan	927	27	100.0	184	4	ABO70242	AbO70242 Pseudomon
855	27	100.0	166	8	ADT58140	Adt58140 Plant pol	928	27	100.0	184	8	ADX76801	Adx76801 Plant ful
856	27	100.0	167	4	ABBI1904	Abb11904 Human IKK	929	27	100.0	184	8	ADY10473	Ady10473 Plant ful
857	27	100.0	167	4	AAI79952	Aam79952 Human pro	930	27	100.0	184	8	ADY13963	Ady13963 Plant ful
858	27	100.0	167	4	AAO01504	Aao01504 Human pol	931	27	100.0	184	8	ADY04403	Ady04403 Plant ful
859	27	100.0	167	7	ABO79119	AbO79119 Pseudomon	932	27	100.0	185	2	AAW26680	Aaw26680 Mature mo
860	27	100.0	167	7	ABO81600	AbO81600 Pseudomon	933	27	100.0	185	2	AAI16692	Aai16692 WO9914235
861	27	100.0	167	7	ABO80182	AbO80182 Pseudomon	934	27	100.0	185	7	ADC08149	Adc08149 Rice prot
862	27	100.0	167	8	ABO54950	AbO54950 Human gen	935	27	100.0	185	7	ABO81857	AbO81857 Pseudomon
863	27	100.0	168	3	AAI68732	Aay68732 Amino aci	936	27	100.0	185	7	ABO71939	AbO71939 Pseudomon
864	27	100.0	169	7	ABO76659	AbO76659 Pseudomon	937	27	100.0	185	7	ABM87894	Abm87894 Rice abio
865	27	100.0	170	7	ABO77560	AbO77560 Pseudomon	938	27	100.0	186	4	AAI83887	Aai83887 Human imm
866	27	100.0	170	8	ADG62476	Ado62476 Transcrip	939	27	100.0	186	7	ABO70923	AbO70923 Pseudomon
867	27	100.0	171	3	AGI11967	Aag11967 Arabidops	940	27	100.0	187	4	AAE03928	Aae03928 Human gen
868	27	100.0	171	3	AGA49813	Aag49813 Arabidops	941	27	100.0	187	5	ABE57411	AbE57411 Human sec
869	27	100.0	171	7	ABO61718	AbO61718 Klebsiell	942	27	100.0	187	6	ABR01052	AbR01052 Human gen
870	27	100.0	171	8	ADY05085	Ady05085 Plant ful	943	27	100.0	187	6	ADA44070	Ada44070 Human sec
871	27	100.0	172	2	AAI29484	Aay29484 Rat CIRP.	944	27	100.0	187	6	ABM87619	Abm87619 Rice abio
872	27	100.0	172	3	ABA42490	Aab42490 Human ORF	945	27	100.0	187	8	ADT60398	Adt60398 Plant pol
873	27	100.0	172	4	AAI83884	Aam83884 Human imm	946	27	100.0	187	8	ADT50893	Adt50893 Cancer re
874	27	100.0	172	6	ABO53056	AbO53056 Human put	947	27	100.0	187	8	ADX79961	Adx79961 Plant ful
875	27	100.0	172	7	ABM89500	Abm89500 Rice abio	948	27	100.0	188	3	AAI51647	Aai51647 Human sec
876	27	100.0	172	8	ADX68887	Adx68887 Plant ful	949	27	100.0	188	4	AAU01739	Aau01739 Human sec
877	27	100.0	172	9	ADY14288	Ady14288 PRO polyp	950	27	100.0	188	4	AAU01638	Aau01638 Human sec
878	27	100.0	172	9	ADY14286	Ady14286 PRO polyp	951	27	100.0	188	5	ABG91413	Abg91413 Primate L
879	27	100.0	173	2	AAW45092	Aaw45092 Residues	952	27	100.0	188	7	ABO77619	AbO77619 Pseudomon
880	27	100.0	174	3	AGI38222	Aag38222 Arabidops	953	27	100.0	188	8	ADX80486	Adx80486 Plant ful
881	27	100.0	174	3	AGI12956	Aag12956 Arabidops	954	27	100.0	189	8	ADQ16270	Adq16270 Thalesres
882	27	100.0	174	4	ABBA1717	Abba1717 Peptide #	955	27	100.0	189	9	ADY10082	Ady10082 Plant ful
883	27	100.0	174	4	AAI75401	Aam75401 Human bon	956	27	100.0	189	9	ADZ00608	Adz00608 G3045 pol
884	27	100.0	174	4	AAI62591	Aam62591 Human bra	957	27	100.0	190	8	ADX68384	Adx68384 Plant ful
885	27	100.0	175	9	ADY65788	Ady65788 S. mansoni	958	27	100.0	191	7	ABM87043	Abm87043 Rice abio
886	27	100.0	176	2	AAW45093	Aaw45093 Residues	959	27	100.0	191	8	ADT58473	Adt58473 Plant pol
887	27	100.0	176	3	AGI12955	Aag12955 Arabidops	960	27	100.0	191	8	ADY10244	Ady10244 Plant ful
888	27	100.0	176	3	AGI38221	Aag38221 Arabidops	961	27	100.0	191	8	ADY14559	Ady14559 Plant ful
889	27	100.0	176	4	ABG14276	Abg14276 Novel hum	962	27	100.0	193	8	ADY12691	Ady12691 Plant ful
890	27	100.0	176	8	ADX79831	Adx79831 Plant ful	963	27	100.0	193	8	ADX76700	Adx76700 Plant ful
891	27	100.0	176	8	ADY24608	Ady24608 Plant ful	964	27	100.0	193	8	ADX91599	Adx91599 Plant ful
892	27	100.0	176	8	ADX60182	Adx60182 Plant ful	965	27	100.0	194	2	AAW72221	Aaw72221 HSV-2 str
893	27	100.0	176	9	AEA13749	Aea13749 Human alp	966	27	100.0	194	4	ABG04824	Abg04824 Novel hum
894	27	100.0	177	5	AAU85542	Aau85542 cDNA enco	967	27	100.0	194	7	ABM87026	Abm87026 Rice abio
895	27	100.0	177	5	ABU67203	Abu67203 G-protein	968	27	100.0	194	7	ABM86946	Abm86946 Rice abio
896	27	100.0	177	6	ABP53984	Abp53984 Human adr	969	27	100.0	194	7	ABM87196	Abm87196 Rice abio
897	27	100.0	177	6	ABU69514	Abu69514 Human lun	970	27	100.0	195	4	AAE01621	Aae01621 Human gen
898	27	100.0	177	6	ABU66417	Abu66417 Lung canc	971	27	100.0	195	5	ABG63792	Abg63792 Human alb
899	27	100.0	177	7	ADH47197	Adh47197 Human lun	972	27	100.0	195	5	ABF51968	Abf51968 Human RNA
900	27	100.0	177	7	ABO78107	AbO78107 Pseudomon	973	27	100.0	195	8	ADL77057	Adl77057 Albumin E

974	27	100.0	195	8	ADT59438	Plant pol
975	27	100.0	196	2	AAW73911	Mutant E2
976	27	100.0	196	7	ABM89487	Rice abio
977	27	100.0	196	7	ABM89950	Rice abio
978	27	100.0	196	8	ABO58595	Human gen
979	27	100.0	196	8	ADx71760	Human gen
980	27	100.0	196	8	ADx72013	Plant ful
981	27	100.0	196	8	ADy04777	Plant ful
982	27	100.0	198	3	AAg53458	Arabidops
983	27	100.0	198	3	AAg23227	Arabidops
984	27	100.0	198	4	AAb95499	Human pro
985	27	100.0	198	7	ABO78641	Pseudomon
986	27	100.0	199	4	AAb82388	Cell cycl
987	27	100.0	201	6	ABp77335	N. gonorr
988	27	100.0	201	6	ABu37797	Protein e
989	27	100.0	201	6	ABu37400	Protein e
990	27	100.0	201	8	ADP08355	Neisseria
991	27	100.0	201	9	ABE93465	Neisseria
992	27	100.0	202	3	AAg23226	Arabidops
993	27	100.0	202	3	AAg53457	Arabidops
994	27	100.0	202	4	AAE03969	Human gen
995	27	100.0	202	4	AAb66004	Human sec
996	27	100.0	202	4	AAb66003	Human sec
997	27	100.0	202	4	AAb75572	Human sec
998	27	100.0	202	4	AAb75571	Gene 19 h
999	27	100.0	202	8	ADn73367	Thale cre
1000	27	100.0	202	8	ADT57401	Plant pol

ALIGNMENTS

RESULT 1					
ID	AAW65560	standard; peptide; 7 AA.			
XX	AAW65560;				
AC	AAW65560;				
XX	27-AUG-2003 (revised)				
DT	15-OCT-1998	(first entry)			
XX	Epstein-Barr virus derived peptide #3.				
DE	Epstein-Barr virus; diagnostic test; autoimmune disease; vaccine;				
XX	infection; antibody; screening; genetic marker.				
KW	Synthetic.				
OS	Human herpesvirus 4.				
XX	WO9830586-A2.				
PN	16-JUL-1998.				
PD	13-JAN-1998;	98WO-US000342.			
PF	13-JAN-1997;	97US-00781296.			
XX	(OKLA-) OKLAHOMA MEDICAL RES FOUND.				
PA	Harley JB, James JA;				
PI	WPI; 1998-399062/34.				
XX	Use of Epstein-Barr virus or component(s) - for developing product(s)				
XX	which can be used for preventing, diagnosing, treating or determining				
PT	risk of developing autoimmune disease.				
XX	Claim 8; Page 64; 81pp; English.				
PS	The invention relates to a vaccine for alleviating or preventing				
XX	autoimmune disorders induced by infection with Epstein-Barr virus (EBV).				
CC	It comprises EBV or a component in a carrier for administration of the				
CC	virus or viral component to alleviate or prevent the autoimmune disorder.				

CC	Also claimed are: (1) a diagnostic test kit comprising: (a) reagents
CC	which can be used to detect levels of antibodies to EBV, indicators of
CC	EBV infection of cells, or levels of EBV DNA or protein in a patient; (b)
CC	control samples from individuals not at risk of developing an autoimmune
CC	disease; and (c) a device for determining the differences in levels of a
CC	patient and control samples to distinguish individuals at higher risk of
CC	developing an autoimmune disease from those at lower risk of developing
CC	an autoimmune disease; and (2) a method for screening for genetic markers
CC	or risk factors for development of autoimmune disorders induced by
CC	infection with EBV comprising comparing the responses of different
CC	strains of the same species of an animal vaccinated with EBV or a
CC	component to induce an autoimmune response in at least one of the strains
CC	and comparing the differences in the genetics of the different strains to
CC	identify potential genetic markers or risk factors. The methods can be
CC	used for the prevention, diagnosis, and treatment of autoimmune diseases
CC	having EBV as an etiological agent. The autoimmune diseases may be e.g.
CC	systemic lupus erythematosus, Sjogren's syndrome, rheumatoid arthritis,
CC	juvenile onset diabetes mellitus, Wegener's granulomatosis, etc. The
CC	present sequence represents a peptide derived from Epstein-Barr virus.
CC	Reagents are used to detect antibodies to this peptide in a specifically
CC	claimed diagnostic test. (Updated on 27-AUG-2003 to correct OS field.)
XX	
XX	Sequence 7 AA;
SQ	
	Query Match 100.0%; Score 27; DB 2; Length 7;
	Best Local Similarity 100.0%; Pred. No. 2e+06;
	Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 RGRGR 5
Db	1 RGRGR 5
RESULT 2	
AAE09064	AAE09064 standard; peptide; 7 AA.
ID	AAE09064 standard; peptide; 7 AA.
XX	AAE09064;
AC	AAE09064;
XX	11-SEP-2003 (revised)
DT	15-NOV-2001 (first entry)
XX	Epstein-Barr virus (EBV) nuclear antigen-1 (EBNA-1) peptide #3.
XX	Vaccine; therapy; autoimmune disorder; Epstein-Barr virus; EBV; anaemia;
KW	polymyositis; systemic lupus erythematosus; SLE; rheumatoid arthritis;
KW	Sjogren's syndrome; diabetes mellitus; adenitis; multiple sclerosis;
KW	demyelinating disease; Hashimoto's thyroiditis; autoimmune infertility;
KW	hypoparathyroidism; primary biliary cirrhosis; ankylosing spondylitis;
KW	inflammatory bowel disease; Addison's disease; thyroiditis; filariasis;
KW	Graves' disease; ulcerative colitis; dermatomyositis; myasthenia gravis;
KW	Crohn's disease; CREST syndrome; autoimmune cystitis; glomerulonephritis;
KW	polyarteritis nodosa; hepatitis; atopic rhinitis; Goodpasture's syndrome;
KW	sarcoidosis; rheumatic fever; anti-phospholipid syndrome; farmer's lung;
KW	Cushing's syndrome; bird-fancier's lung; alveolitis; erythema nodosum;
KW	pyoderma gangrenosum; fibromyalgia; Kawasaki's disease; cardiomyopathy;
KW	Sampner's syndrome; asthma; polymyalgia rheumatica; psoriasis; arthritis;
KW	erythroblastosis foetalis; cycloitis; IGA nephropathy; Hodgkin's lymphoma;
KW	renal cell carcinoma; eosinophilia; immunosuppressive; ophthalmological;
KW	thrombimetic; neuroprotective; cytostatic; nephrotropic; antiallergic;
KW	dengue; antiulcer; vasotropic; antipyretic; hepatotropic.
XX	
OS	Human herpesvirus 4.
XX	
XX	WO200158481-A2.
PN	
PD	16-AUG-2001.
XX	
XX	09-FEB-2001; 2001WO-US004191.
PF	
XX	09-FEB-2000; 2000US-00500904.
PR	
XX	(OKLA-) OKLAHOMA MEDICAL RES FOUND.
PA	

```

XX Harley JB, James JA, Kaufman KM;
XX WPI; 2001-522437/57.
XX Novel vaccine for alleviating or preventing autoimmune disorders induced
XX Epstein-Barr virus (EBV) infection e.g. systemic lupus erythematosus,
XX juvenile onset diabetes mellitus, comprises EBV virus or its component.
XX Claim 4; Page 60; 114pp; English.
XX The present invention relates to a vaccine for alleviating or preventing
XX autoimmune disorders induced by infection with Epstein-Barr virus (EBV),
XX comprising EBV or its component in a carrier. The vaccine is useful for
XX preventing or alleviating autoimmune disorders induced by EBV, e.g.
XX systemic lupus erythematosus (SLE), Sjogren's syndrome, juvenile onset
XX diabetes mellitus, rheumatoid arthritis, Wegener's granulomatosis,
XX inflammatory bowel disease, polymyositis, dermatomyositis, multiple
XX endocrine failure, Schmidt's syndrome, autoimmune uveitis, Addison's
XX disease, adrenailitis, primary biliary cirrhosis, Graves' disease,
XX thyroiditis, Hashimoto's thyroiditis, autoimmune thyroid disease,
XX pernicious and haemolytic anaemia, lupoid hepatitis, demyelinating
XX disease, multiple sclerosis, subacute cutaneous lupus erythematosus,
XX hypoparathyroidism, Dressler's syndrome, myasthenia gravis, autoimmune
XX idiopathic thrombocytopenic purpura, autoimmune pemphigus vulgaris,
XX pemphigus, bullous pemphigoid, dermatitis herpetiformis, alopecia areata,
XX autoimmune cystitis, pemphigoid, scleroderma, progressive systemic
XX sclerosis, CREST syndrome (Calcinosis, Raynaud's oesophageal dysmotility,
XX sclerodactyly and telangiectasis), adult onset diabetes mellitus (Type II
XX diabetes), male or female autoimmune infertility, ankylosing spondylitis,
XX ulcerative colitis, Crohn's disease, mixed connective tissue disease,
XX polyarteritis nodosa, systemic necrotising vasculitis,
XX glomerulonephritis, atopic dermatitis, atopic rhinitis, Goodpasture's
XX syndrome, Chagas' disease, sarcoidosis, rheumatic fever, asthma,
XX recurrent abortion, anti-phospholipid syndrome, farmer's lung, erythema
XX multiforme, postcardotomy syndrome, Cushing's syndrome, autoimmune
XX chronic active hepatitis, bird-fancier's lung, allergic
XX encephalomyelitis, toxic necrodermal lysis, alopecia, Alport's syndrome,
XX allergic alveolitis, fibrosing alveolitis, interstitial lung disease,
XX erythema nodosum, pyoderma gangrenosum, transfusion reaction, chronic
XX fatigue syndrome, fibromyalgia, Takayasu's arteritis, Kawasaki's disease,
XX polynyalgia rheumatica, temporal arteritis, giant cell arteritis, dengue,
XX Samper's syndrome (triaditis, nasal polyps, eosinophilia) and Behcet's
XX disease, Caplan's syndrome, encephalomyositis, erythema elevatum et
XX diutinum, psoriasis, erythroblastosis foetalis, Shulman's syndrome, IGA
XX nephropathy, Felty's syndrome, fascitis with eosinophilia, filariasis,
XX chronic cystitis, heterochromic cyclitis, Fuch's cyclitis, Hodgkin's and
XX non-Hodgkin's lymphoma, cardiomyopathy, Henoch-Schonlein purpura, post
XX vaccination syndromes, renal cell carcinoma, Eaton-Lambert syndrome or
XX relapsing polychondritis. The present sequence is Epstein-Barr virus
XX (EBV) nuclear antigen-1 (EBNA-1) peptide. (Updated on 11-SEP-2003 to
XX standardise OS field)
XX Sequence 7 AA;

Query Match 100.0%; Score 27; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 2e+06;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
Db 1 RGRGR 5

RESULT 3
ABU07629
ID ABU07629 standard; peptide; 7 AA.
XX ABU07629;
AC ABU07629;
XX 23-OCT-2003 (revised)
DT 10-MAY-2003 (first entry)
XX

```

---

```

DE Harley JB, James JA, Kaufman KM;
XX WPI; 2001-522437/57.
XX Novel vaccine for alleviating or preventing autoimmune disorders induced
XX Epstein-Barr virus (EBV) infection e.g. systemic lupus erythematosus,
XX juvenile onset diabetes mellitus, comprises EBV virus or its component.
XX Claim 4; Page 60; 114pp; English.
XX The present invention relates to a vaccine for alleviating or preventing
XX autoimmune disorders induced by infection with Epstein-Barr virus (EBV),
XX comprising EBV or its component in a carrier. The vaccine is useful for
XX preventing or alleviating autoimmune disorders induced by EBV, e.g.
XX systemic lupus erythematosus (SLE), Sjogren's syndrome, juvenile onset
XX diabetes mellitus, rheumatoid arthritis, Wegener's granulomatosis,
XX inflammatory bowel disease, polymyositis, dermatomyositis, multiple
XX endocrine failure, Schmidt's syndrome, autoimmune uveitis, Addison's
XX disease, adrenailitis, primary biliary cirrhosis, Graves' disease,
XX thyroiditis, Hashimoto's thyroiditis, autoimmune thyroid disease,
XX pernicious and haemolytic anaemia, lupoid hepatitis, demyelinating
XX disease, multiple sclerosis, subacute cutaneous lupus erythematosus,
XX hypoparathyroidism, Dressler's syndrome, myasthenia gravis, autoimmune
XX idiopathic thrombocytopenic purpura, autoimmune pemphigus vulgaris,
XX pemphigus, bullous pemphigoid, dermatitis herpetiformis, alopecia areata,
XX autoimmune cystitis, pemphigoid, scleroderma, progressive systemic
XX sclerosis, CREST syndrome (Calcinosis, Raynaud's oesophageal dysmotility,
XX sclerodactyly and telangiectasis), adult onset diabetes mellitus (Type II
XX diabetes), male or female autoimmune infertility, ankylosing spondylitis,
XX ulcerative colitis, Crohn's disease, mixed connective tissue disease,
XX polyarteritis nodosa, systemic necrotising vasculitis,
XX glomerulonephritis, atopic dermatitis, atopic rhinitis, Goodpasture's
XX syndrome, Chagas' disease, sarcoidosis, rheumatic fever, asthma,
XX recurrent abortion, anti-phospholipid syndrome, farmer's lung, erythema
XX multiforme, postcardotomy syndrome, Cushing's syndrome, autoimmune
XX chronic active hepatitis, bird-fancier's lung, allergic
XX encephalomyelitis, toxic necrodermal lysis, alopecia, Alport's syndrome,
XX allergic alveolitis, fibrosing alveolitis, interstitial lung disease,
XX erythema nodosum, pyoderma gangrenosum, transfusion reaction, chronic
XX fatigue syndrome, fibromyalgia, Takayasu's arteritis, Kawasaki's disease,
XX polynyalgia rheumatica, temporal arteritis, giant cell arteritis, dengue,
XX Samper's syndrome (triaditis, nasal polyps, eosinophilia) and Behcet's
XX disease, Caplan's syndrome, encephalomyositis, erythema elevatum et
XX diutinum, psoriasis, erythroblastosis foetalis, Shulman's syndrome, IGA
XX nephropathy, Felty's syndrome, fascitis with eosinophilia, filariasis,
XX chronic cystitis, heterochromic cyclitis, Fuch's cyclitis, Hodgkin's and
XX non-Hodgkin's lymphoma, cardiomyopathy, Henoch-Schonlein purpura, post
XX vaccination syndromes, renal cell carcinoma, Eaton-Lambert syndrome or
XX relapsing polychondritis. The present sequence is Epstein-Barr virus
XX (EBV) nuclear antigen-1 (EBNA-1) peptide. (Updated on 11-SEP-2003 to
XX standardise OS field)
XX Sequence 7 AA;

Query Match 100.0%; Score 27; DB 6; Length 7;
Best Local Similarity 100.0%; Pred. No. 2e+06;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
Db 1 RGRGR 5

RESULT 4
AD120905
ID AD120905 standard; peptide; 7 AA.
XX

```

---

```

DE Epstein-Barr virus nuclear antigen peptide #3.
XX EBV; viral; Epstein-Barr virus nuclear antigen; vaccine;
XX autoimmune disorder; Epstein-Barr virus infection; diabetes mellitus;
XX arthritis; multiple sclerosis; dermatitis; psoriasis; asthma; anaemia;
XX allergy.
XX Human herpesvirus 4.
XX OS
XX US2002164355-A1.
XX PN
XX 07-NOV-2002.
XX PD
XX 24-OCT-2001; 2001US-00012756.
XX PF
XX 30-NOV-1993; 93US-00160604.
XX PR 16-MAY-1996; 96US-0019053P.
XX PR 13-JAN-1997; 97US-00781296.
XX PA (HARL/) HARLEY J B.
XX PA (JAME/) JAMES J A.
XX PI Harley JB, James JA;
XX PI WPI; 2003-298686/29.
XX DR
XX New vaccine preventing or alleviating autoimmune disorders induced by the
XX Epstein-Barr virus, such as diabetes mellitus, rheumatoid arthritis,
XX multiple sclerosis, systemic lupus erythematosus, atopic dermatitis and
XX psoriasis.
XX Claim 8; Page 28; 41pp; English.
XX The invention relates to a vaccine for alleviating or preventing
XX autoimmune disorders induced by infection with Epstein-Barr virus,
XX comprising an Epstein-Barr virus or a component in a carrier for,
XX administration to alleviate or prevent the autoimmune disorders. The
XX methods and compositions of the present invention are useful for
XX diagnosing, preventing, treating and/or alleviating autoimmune disorders,
XX such as diabetes mellitus, rheumatoid arthritis, juvenile rheumatoid
XX arthritis, osteoarthritis, psoriatic arthritis, multiple sclerosis,
XX encephalomyelitis, myasthenia gravis, systemic lupus erythematosus,
XX autoimmune thyroiditis, atopic dermatitis, eczematous dermatitis,
XX psoriasis, Sjogren's Syndrome, Crohn's disease, aphthous ulcer, iritis,
XX conjunctivitis, keratoconjunctivitis, ulcerative colitis, asthma,
XX allergic asthma, cutaneous lupus erythematosus, scleroderma, vaginitis,
XX proctitis, drug eruptions, leprosy reversal reactions, erythema
XX nodosum, leprosy, autoimmune uveitis, allergic encephalomyelitis, acute
XX necrotising haemorrhagic encephalopathy, idiopathic bilateral progressive
XX sensorineural hearing loss, aplastic anaemia, pure red cell anaemia,
XX idiopathic thrombocytopenia, polychondritis, Wegener's granulomatosis,
XX chronic active hepatitis, Stevens-Johnson syndrome, idiopathic sprue,
XX lichen planus, Graves' disease, sarcoidosis, primary biliary cirrhosis,
XX uveitis posterior, interstitial lung fibrosis, graft-versus-host disease,
XX and allergy. The present sequence represents an Epstein-Barr virus
XX nuclear antigen peptide used in the method of the invention. (Updated on
XX 23-OCT-2003 to standardise OS field)
XX Sequence 7 AA;

Query Match 100.0%; Score 27; DB 6; Length 7;
Best Local Similarity 100.0%; Pred. No. 2e+06;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
Db 1 RGRGR 5

RESULT 4
AD120905
ID AD120905 standard; peptide; 7 AA.
XX

```

AC ADI20905;  
XX 22-APR-2004 (first entry)  
XX Arginine containing peptide #7 of the invention.  
DE Arginine containing peptide; cholecystokinin; obesity; hyperphagia;  
KW hyperphagia.  
KW Unidentified.  
XX JP2004010569-A.  
XX 15-JAN-2004.  
XX 10-JUN-2002; 2002JP-00168694.  
XX 10-JUN-2002; 2002JP-00168694.  
XX (ASAN/) ASANO Y.  
XX WPI; 2004-160930/16.  
XX New arginine containing peptide has cholecystokinin secretion promoting  
PT activity, useful in foodstuffs for preventing obesity or hyperphagia.  
XX Example 1; SEQ ID NO 7; 12pp; Japanese.  
XX The present invention relates to an arginine containing peptide has  
CC cholecystokinin secretion promoting activity. The method has anorectic  
CC effect is useful for preventing obesity or hyperphagia and has efficient  
CC cholecystokinin secretion promotion activity. The method is cheaper, and  
CC safe in preventing obesity or hyperphagia. The present sequence  
CC represents an arginine containing peptide of the invention.  
XX Sequence 7 AA;  
SQ  
Query Match 100.0%; Score 27; DB 8; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RGRGR 5  
DB 2 RGRGR 6  
RESULT 5  
AD017054  
ID AD017054 standard; peptide; 7 AA.  
AC AD017054;  
XX 29-JUL-2004 (first entry)  
XX Epstein-Barr virus nuclear antigen 1, antigenic peptide #2.  
DE vaccine; autoimmune disorder; Epstein-Barr virus; EBV;  
KW systemic lupus erythematosus; Sjogren's syndrome; rheumatoid arthritis;  
KW juvenile onset diabetes mellitus; Wegener's granulomatosis;  
KW inflammatory bowel disease; Epstein-Barr nuclear antigen 1.  
XX Human herpesvirus 4.  
XX US2004086522-A1.  
XX 06-MAY-2004.  
XX 27-JUN-2003; 2003US-00607918.  
XX 30-NOV-1993; 93US-00160604.  
PR 16-MAY-1996; 96US-0019053P.  
PR 13-JAN-1997; 97US-00781296.  
PR 24-OCT-2001; 2001US-00012756.

XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
XX Harley JB, James JA;  
XX WPI; 2004-356164/33.  
XX New vaccine for alleviating or preventing autoimmune disorders induced by  
PT Epstein-Barr virus (EBV) infection, e.g. systemic lupus erythematosus,  
PT comprises EBV or its component in a pharmaceutical carrier.  
XX Claim 8; Page 17; 30pp; English.  
XX The invention relates to a new vaccine for alleviating or preventing  
CC autoimmune disorders induced by infection with Epstein-Barr virus (EBV).  
CC The vaccine comprises EBV or its component in a pharmaceutical carrier  
CC for administration of the virus or viral component in an amount and mode  
CC of administration to alleviate or prevent the autoimmune disorders. The  
CC composition and methods are useful for diagnosing, preventing or treating  
CC autoimmune diseases caused by Epstein-Barr virus, such as systemic lupus  
CC erythematosus, Sjogren's syndrome, rheumatoid arthritis, juvenile onset  
CC diabetes mellitus, Wegener's granulomatosis or inflammatory bowel  
CC disease. These may also be used in screening of therapeutics for  
CC prevention or alleviation of autoimmune disorders induced by EBV  
CC infection. The present sequence represents an Epstein-Barr virus nuclear  
CC antigen 1, antigenic peptide used to make the vaccine of the invention.  
XX Sequence 7 AA;  
SQ  
Query Match 100.0%; Score 27; DB 8; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RGRGR 5  
DB 1 RGRGR 5  
RESULT 6  
AAW65559  
ID AAW65559 standard; peptide; 8 AA.  
XX AAW65559;  
AC AAW65559;  
XX 27-AUG-2003 (revised)  
DT 15-OCT-1998 (first entry)  
XX Epstein-Barr virus derived peptide #2.  
DE Epstein-Barr virus; diagnostic test; autoimmune disease; vaccine;  
XX infection; antibody; screening; genetic marker.  
KW Synthetic.  
XX Human herpesvirus 4.  
OS WO9830586-A2.  
XX 16-JUL-1998.  
PD 13-JAN-1998; 98WO-US0000342.  
XX 13-JAN-1997; 97US-00781296.  
PR (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
XX Harley JB, James JA;  
XX WPI; 1998-399062/34.  
XX Use of Epstein-Barr virus or component (s) - for developing product(s)  
PT which can be used for preventing, diagnosing, treating or determining  
PT risk of developing autoimmune disease.  
XX



XX AC ABG75621;  
XX DT 25-APR-2003 (first entry)  
XX DE Synthetic peptide for protein ligation #2.  
XX KW Protein ligation; intein-chitin binding domain; CBD; N-terminal cysteine;  
KW unoxidised sulphhydryl side chain; protein-chip; semi-synthetic protein;  
KW abelson protein tyrosine kinase SH3 domain.  
XX OS Synthetic.  
XX FH Key Location/Qualifiers  
FT Modified-site 8  
FT /label= OTHER  
FT /note= "Lysine is covalently attached to a fluorescein  
FT moiety"  
XX US2002151006-A1.  
XX PD 17-OCT-2002.  
XX PF 12-JUL-2001; 2001US-00904117.  
XX PR 13-NOV-1997; 97US-0065391P.  
PR 24-JUL-1998; 98US-0093990P.  
PR 13-NOV-1998; 98US-00191890.  
XX PA (MUIR/) MUIR T W.  
PA (COLE/) COLE P A.  
PA (FRIE/) FRIEDMAN J M.  
PA (SONDH/) SONDHI D.  
PA (SEVE/) SEVERINOV K.  
XX PI Muir TW, Cole PA, Friedman JM, Sondhi D, Severinov K;  
XX WPI; 2003-238034/23.  
XX DR  
XX PT Cleaving recombinantly expressed proteins bound to intein-chitin binding  
PT domains, and ligating to peptides with N-terminal cysteines having  
PT unoxidized sulphhydryl side chain, useful for producing semi-synthetic  
PT proteins.  
XX PS Example 3; Page 14; 35pp; English.  
XX CC The invention relates to cleaving a recombinantly expressed protein bound  
CC to an intein-chitin binding domain (CBD) and ligating the cleaved protein  
CC to a peptide with an N-terminal cysteine having an unoxidised sulphhydryl  
CC side chain, by contacting the bound protein with a peptide in the  
CC presence of conjugated thiol to effect cleavage of the protein from the  
CC intein-CBD and production of a C-terminal thioester of the protein which  
CC undergoes intramolecular rearrangement to form amide bond linking the  
CC protein to the peptide. The method is used for linking recombinantly  
CC expressed protein bound to an intein-CBD to a peptide with N-terminal  
CC cysteine having an unoxidised sulphhydryl side chain. The method is  
CC useful for producing a protein-chip which is useful for diagnostic  
CC screening for a specific protein, antibody or antigen and for preparing  
CC semi-synthetic proteins of any size. The protein-chip is also useful for  
CC identifying presence of protein in a sample. The method is also useful  
CC for generating recombinant proteins e.g. cytotoxic recombinant proteins,  
CC recombinant proteins partially labeled with a detectable marker, in  
CC nuclear magnetic resonance (NMR) spectroscopy, where proteins used in the  
CC method are segmentally labeled, and for generating semi-synthetic  
CC proteins to facilitate two recombinant, folded proteins to be ligated  
CC together. The present sequence represents a synthetic peptide with an N-  
CC terminal cysteine which was ligated to an intein-CBD-immobilised human  
CC abelson protein tyrosine kinase SH3 domain  
XX SQ Sequence 8 AA;  
Query Match 100.0%; Score 27; DB 6; Length 8;  
Best Local Similarity 100.0%; Pred. No. 2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RGRGR 5  
Db 3 RGRGR 7  
RESULT 9  
ABB82936  
ID ABB82936 standard; peptide; 8 AA.  
XX AC ABB82936;  
XX DT 14-APR-2003 (first entry)  
XX DE Arginie-dimethylated synthetic peptide seq Id No. 6.  
XX KW Dimethylarginine; mass spectroscopy; monomethylamine; dimethylamine; DMA;  
KW MMA; dimethylcarbodiimide; DMC.  
XX OS Synthetic.  
XX PN WO200294777-A2.  
XX PD 28-NOV-2002.  
XX PF 20-MAY-2002; 2002WO-US015613.  
XX PR 18-MAY-2001; 2001US-0292075P.  
XX PA (MDSP-) MDS PROTEOMICS INC.  
XX PI Brane CJ, Mcbroom LDB;  
XX WPI; 2003-140342/13.  
XX PT Identifying structure of dimethyl arginine for proteomics business by  
PT obtaining neutral loss spectra of peptide containing dimethylarginine  
PT residues by mass spectrometry.  
XX PS Example 1; Page 29; 45pp; English.  
XX CC The invention relates to identifying the structure of dimethylarginine.  
CC The method involves (i) obtaining a neutral loss spectrum of a peptide  
CC containing a dimethylarginine by mass spectroscopy and (ii) determining  
CC if the neutral loss spectrum shows one or both of neutral loss of  
CC monomethylamine (MMA), dimethylcarbodiimide (DMC) and/or neutral loss of  
CC dimethylamine (DMA). The neutral loss of MMA and DMC indicates the  
CC presence of a symmetrically dimethylated arginine residue and the neutral  
CC loss of (DMA) indicates the presence of asymmetrically dimethylated  
CC arginine residue. Sequences ABB82931-939 represent synthetic peptides  
CC having symmetrically and asymmetrically dimethylated arginine residues  
CC used in an experiment to determine if mass spectrometry could be used  
CC to distinguish the peptides that have symmetrically versus asymmetrically  
CC dimethylated arginine residues  
XX SQ Sequence 8 AA;  
Query Match 100.0%; Score 27; DB 6; Length 8;  
Best Local Similarity 100.0%; Pred. No. 2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RGRGR 5  
Db 4 RGRGR 8  
RESULT 10  
ABU07628  
ID ABU07628 standard; peptide; 8 AA.  
XX AC ABU07628;  
XX SQ Sequence 8 AA;  
Query Match 100.0%; Score 27; DB 6; Length 8;  
Best Local Similarity 100.0%; Pred. No. 2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



DT		23-OCT-2003	(revised)
DD		10-MAY-2003	(first entry)
XX			
DE	X	Epstein-Barr virus nuclear antigen peptide #2.	
XX			
KW	X	EBV; viral; Epstein-Barr virus nuclear antigen; vaccine;	
KW	X	autoimmune disorder; Epstein-Barr virus infection; diabetes mellitus;	
KW	X	arthritis; multiple sclerosis; dermatitis; psoriasis; asthma; anaemia;	
KX	X	allergy.	
OS	X		
XS		Human herpesvirus 4.	
PB	N	US2002164355-A1.	
PD	X		
PP	X	07-NOV-2002.	
PR	X	24-OCT-2001; 2001US-00012756.	
PR	X	30-NOV-1993; 93US-00160604.	
PR	X	18-MAY-1996; 96US-0019053P.	
PR	X	13-JAN-1997; 97US-00781296.	
PA	X	(HARLEY) HARLEY J B.	
PI	X	(JAMES/) JAMES J A.	
PI	X	Harley JB, James JA;	
PS	X	WPI; 2003-298686/29.	
DR	X	New vaccine preventing or alleviating autoimmune disorders induced by the	
PT	X	Epstein-Barr virus, such as diabetes mellitus, rheumatoid arthritis,	
PT	X	multiple sclerosis, systemic lupus erythematosus, atopic dermatitis and	
PT	X	psoriasis.	
XX			
CS		Claim 8; Page 28; 4lpp; English.	
CC		The invention relates to a vaccine for alleviating or preventing	
CC		autoimmune disorders induced by infection with Epstein-Barr virus,	
CC		comprising an Epstein-Barr virus or a component in a carrier for	
CC		administration to alleviate or prevent the autoimmune disorders. The	
CC		methods and compositions of the present invention are useful for	
CC		diagnosing, preventing, treating and/or alleviating autoimmune disorders,	
CC		such as diabetes mellitus, rheumatoid arthritis, juvenile rheumatoid	
CC		arthritis, osteoarthritis, psoriatic arthritis, multiple sclerosis,	
CC		encephalomyelitis, myasthenia gravis, systemic lupus erythematosus,	
CC		autoimmune thyroiditis, atopic dermatitis, eczematous dermatitis,	
CC		psoriasis, Sjogren's Syndrome, Crohn's disease, aphthous ulcer, iritis,	
CC		conjunctivitis, keratoconjunctivitis, ulcerative colitis, asthma,	
CC		allergic asthma, cutaneous lupus erythematosus, scleroderma, vaginitis,	
CC		proctitis, drug eruptions, leprosy reversal reactions, erythema	
CC		nodosumlepraem, autoimmune uveitis, allergic encephalomyelitis, acute	
CC		necrotising haemorrhagic encephalopathy, idiopathic bilateral progressive	
CC		sensorineural hearing loss, aplastic anaemia, pure red cell anaemia,	
CC		idiopathic thrombocytopenia, polychondritis, Wegener's granulomatosis,	
CC		chronic active hepatitis, Stevens-Johnson syndrome, idiopathic sprue,	
CC		lichen planus, Graves' disease, sarcoidosis, primary biliary cirrhosis,	
CC		uveitis posterior, interstitial lung fibrosis, graft-versus-host disease,	
CC		and allergy. The present sequence represents an Epstein-Barr virus	
CC		nuclear antigen peptide used in the method of the invention. (Updated on	
CC		23-OCT-2003 to standardise OS field)	
SQ		Sequence 8 AA;	
		Query Match 100.0%; Score 27; DB 6; Length 8;	
		Best Local Similarity 100.0%; Pred. No. 2e+06;	
		Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
QY		1 RGRGR 5	
Dd			
		2 RGRGR 6	
RESULT		11	

KW vaccine; autoimmune disorder; Epstein-Barr virus; EBV;  
KW systemic lupus erythematosus; Sjogren's syndrome; rheumatoid arthritis;  
KW juvenile onset diabetes mellitus; Wegener's granulomatosis;  
KW inflammatory bowel disease; Epstein-Barr nuclear antigen 1.  
OS Human herpesvirus 4.  
XX US2004086522-A1.  
PN 06-MAY-2004.  
XX 27-JUN-2003; 2003US-00607918.  
XX 30-NOV-1993; 93US-00160604.  
PR 16-MAY-1996; 96US-0019053P.  
PR 13-JAN-1997; 97US-00781296.  
PR 24-OCT-2001; 2001US-00012756.  
XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
PA Harley JB, James JA;  
PI WPI; 2004-356164/33.  
DR New vaccine for alleviating or preventing autoimmune disorders induced by  
PT Epstein-Barr virus (EBV) infection, e.g. systemic lupus erythematosus,  
PT comprises EBV or its component in a pharmaceutical carrier.  
XX Claim 8; Page 17; 30pp; English.  
XX The invention relates to a new vaccine for alleviating or preventing  
CC autoimmune disorders induced by infection with Epstein-Barr virus (EBV).  
CC The vaccine comprises EBV or its component in a pharmaceutical carrier  
CC for administration of the virus or viral component in an amount and mode  
CC of administration to alleviate or prevent the autoimmune disorders. The  
CC composition and methods are useful for diagnosing, preventing or treating  
CC autoimmune diseases caused by Epstein-Barr virus, such as systemic lupus  
CC erythematosus, Sjogren's syndrome, rheumatoid arthritis, juvenile onset  
CC diabetes mellitus, Wegener's granulomatosis or inflammatory bowel  
CC disease. These may also be used in screening of therapeutics for  
CC prevention or alleviation of autoimmune disorders induced by EBV  
CC infection. The present sequence represents an Epstein-Barr virus nuclear  
CC antigen 1, antigenic peptide used to make the vaccine of the invention.  
XX  
SQ Sequence 8 AA;  
Query Match 100.0%; Score 27; DB 8; Length 8;  
Best Local Similarity 100.0%; Pred. No. 2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RGRGR 5  
DB |||||  
2 RGRGR 6  
RESULT 13  
ID AAW60714 standard; peptide; 9 AA.  
XX AAW60714;  
AC AAW60714;  
XX 03-SEP-1998 (first entry)  
DT Peptide used in the course of the invention.  
DE Spacer; increase; efficiency; peptide synthesis; hydrophobic peptide;  
KW peptide nucleic acid; PNA.  
XX Synthetic.  
OS W09817677-A1.  
PN

PD 30-APR-1998.  
XX 24-OCT-1997; 97WO-AU000711.  
XX 24-OCT-1996; 96AU-00003240.  
PR (UYQU ) UNIV QUEENSLAND.  
PA Englebrethsen D;  
XX WPI; 1998-261419/23.  
DR New spacer for increasing efficiency of solid phase peptide synthesis -  
XX useful for, e.g. synthesis of strongly hydrophobic peptides and peptide  
PT nucleic acids.  
XX Disclosure; Page 6; 49pp; English.  
XX The present peptide is used in the course of the invention. The  
CC specification describes a new spacer for increasing efficiency of  
CC synthesis of peptides or related compounds. The method is particularly  
CC used for synthesis of strongly hydrophobic peptides or of peptide nucleic  
CC acids (PNA)  
XX  
SQ Sequence 9 AA;  
Query Match 100.0%; Score 27; DB 2; Length 9;  
Best Local Similarity 100.0%; Pred. No. 2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RGRGR 5  
DB |||||  
2 RGRGR 6  
RESULT 14  
ADR21207 standard; peptide; 9 AA.  
ID ADR21207  
XX ADR21207;  
AC 21-OCT-2004 (first entry)  
DT Novel cellular drug delivery method peptide RRG.  
XX  
XX antibacterial; virucide; cytostatic; antitubercular; tuberculostatic;  
KW antileprotic; antiparasitic; fungicide; antisense therapy; gene therapy;  
KW electromagnetic radiation; infectious disease; bacterial disease;  
KW tuberculosis; leprosy; viral disease; fungal disease; parasitic disease;  
KW cancer; siRNA; gene silencing; gene expression; small interfering RNA.  
XX  
XX Synthetic.  
OS WO2004063342-A2.  
XX 29-JUL-2004.  
PD 09-JAN-2004; 2004WO-US000430.  
PF 09-JAN-2003; 2003US-0438778P.  
PR (INVI-) INVITROGEN CORP.  
PA Dalby B, Bennett RP;  
XX WPI; 2004-553730/53.  
DR Delivering a polypeptide to a cell for e.g. treating a disease, comprises  
XX contacting the cell with the polypeptide, nucleic acid, fluorescent  
PT molecule, and/or a cellular delivery molecule, and treating to dissociate  
PT the polypeptide.  
XX  
XX Example 1; SEQ ID NO 6; 165pp; English.  
PS

XX The invention relates to a method of delivering (M1) a polypeptide to a  
 CC cell, by contacting the cell with, in any order or combination, the  
 CC polypeptide, nucleic acid, fluorescent molecule, cellular delivery  
 CC molecule and/or a transfection agent, and treating the cell with a  
 CC treatment that results in the dissociation of the polypeptide from the  
 CC nucleic acid, the fluorescent molecule, or/and the cellular delivery  
 CC molecule. (M1) is useful for delivering a polypeptide to a cell. The  
 CC molecules are useful for treating an individual suffering from a disease  
 CC or disorder and for providing gene therapy to an individual in need where  
 CC the treatment further involves exposing an individual to electromagnetic  
 CC radiation. The diseases treated by the molecules include infectious  
 CC diseases such as bacterial diseases e.g., tuberculosis, leprosy, viral  
 CC diseases, fungal diseases, parasitic diseases, and cancer. This sequence  
 CC represents a peptide used in the method of the invention.  
 XX  
 SQ Sequence 9 AA;  
 Query Match 100.0%; Score 27; DB 8; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 2e+06;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RGRGR 5  
 Db 2 RGRGR 6  
 RESULT 15  
 ADS18175  
 ID ADS18175 standard; peptide; 9 AA.  
 AC ADS18175;  
 DT 30-DEC-2004 (first entry)  
 DE Human Sp35 peptide fragment SeqID7.  
 XX  
 KW Sp35; CNS-Gen; neuroprotective; anticonvulsant; nootropic;  
 KW antiparkinsonian; antidiabetic; cerebroprotective; vasotropic;  
 KW gene therapy; CNS disease; neurodegenerative disease; multiple sclerosis;  
 KW Amyotrophic lateral sclerosis; ALS; Huntington's disease;  
 KW Alzheimer's disease; Parkinson's disease; diabetic neuropathy; stroke;  
 KW traumatic brain injury; spinal cord injury; optic nerve injury; human.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2004085648-A2.  
 PD 07-OCT-2004.  
 XX  
 PF 17-MAR-2004; 2004WO-US008323.  
 XX  
 PR 19-MAR-2003; 2003US-0455756P.  
 PR 20-JUN-2003; 2003US-0480241P.  
 PR 01-AUG-2003; 2003US-0492057P.  
 XX  
 PA (BIOG-) BIOGEN IDEC WA INC.  
 XX  
 PI Mi S, McCoy J, Pepinsky RB, Lee DHS;  
 XX  
 XX WPI; 2004-710319/69.  
 DR  
 XX New nucleic acid encoding Sp35 polypeptide, useful for treating a CNS  
 PT disease or injury, e.g. multiple sclerosis, Huntington's disease,  
 PT Parkinson's disease, diabetic neuropathy, stroke, spinal cord injury, or  
 PT optic nerve injury.  
 XX  
 PS Disclosure; SEQ ID NO 7; 70pp; English.  
 XX  
 CC This invention relates to a novel isolated nucleic acid and the Sp35  
 CC protein encoded by it. The invention may be useful for the production of  
 CC compounds with a CNS-gen, neuroprotective, anticonvulsant, nootropic,  
 CC antiparkinsonian, antidiabetic, cerebroprotective or vasotropic activity

CC whilst the disclosed sequences may prove useful for gene therapy. The  
 CC invention may be useful for developing a treatment for a CNS or  
 CC neurodegenerative disease, disorder or injury, for example multiple  
 CC sclerosis, Amyotrophic lateral sclerosis (ALS), Huntington's disease,  
 CC Alzheimer's disease, Parkinson's disease, diabetic neuropathy, stroke,  
 CC traumatic brain injuries, specifically, a spinal cord injury or an optic  
 CC nerve injury. The present sequence is that of a peptide derived from the  
 CC human Sp35 protein of the invention.  
 XX  
 SQ Sequence 9 AA;  
 Query Match 100.0%; Score 27; DB 8; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 2e+06;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RGRGR 5  
 Db 2 RGRGR 6

RESULT 16  
 AAG95106  
 ID AAG95106 standard; peptide; 10 AA.  
 AC AAG95106;  
 DT 18-SEP-2001 (first entry)  
 DE Human complementary peptide, SEQ ID NO: 1300.  
 XX  
 KW Human; complementary peptide; ligand; drug discovery; drug design.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200142277-A2.  
 PD 14-JUN-2001.  
 XX  
 PF 13-DEC-2000; 2000WO-GB004776.  
 XX  
 PR 13-DEC-1999; 99GB-00029464.  
 XX  
 PA (PROT-) PROTEOM LTD.  
 XX  
 PI Roberts GW, Heal JR;  
 XX  
 DR WPI; 2001-408419/43.  
 XX

PT A set of peptide ligands consisting of specific complementary peptides to  
 PT proteins encoded by genes of the human genome, useful in an assay for  
 PT screening and identifying of one or more novel peptides which are drug  
 PT candidates or pro-drugs.  
 XX  
 PS Example 4; Page 226; 646pp; English.  
 XX

CC The invention relates to a set of complementary peptide ligands generated  
 CC from the human genome. The complementary peptides interact with their  
 CC relevant target proteins encoded in the human genome. They can be used as  
 CC reagents in drug discovery and as lead ligands to facilitate drug design  
 CC and development. The present sequence is a complementary peptide provided  
 CC in the specification  
 XX  
 SQ Sequence 10 AA;  
 Query Match 100.0%; Score 27; DB 4; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RGRGR 5  
 Db 2 RGRGR 6



```
CC in the specification
XX
SQ Sequence 10 AA;

Query Match      100.0%; Score 27; DB 4; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 RGRGR 5
Db      |||||
        4 RGRGR 8

RESULT 20
AAG95526
ID AAG95526 standard; peptide; 10 AA.
XX
AC AAG95526;
XX
DT 18-SEP-2001 (first entry)
XX
DE Human complementary peptide, SEQ ID NO: 1720.
XX
KW Human; complementary peptide; ligand; drug discovery; drug design.
XX
OS Homo sapiens.
XX
PN WO200142277-A2.
XX
PD |||||
XX      4 RGRGR 8

RESULT 21
AAG97414
ID AAG97414 standard; peptide; 10 AA.
XX
AC AAG97414;
XX
DT 18-SEP-2001 (first entry)
XX

Human complementary peptide, SEQ ID NO: 3608.
XX
Human; complementary peptide; ligand; drug discovery; drug design.
XX
Homo sapiens.
XX
WO200142277-A2.
XX
14-JUN-2001.
XX
13-DEC-2000; 2000WO-GB004776.
XX
13-DEC-1999; 99GB-00029464.
XX
(PROT-) PROTEOM LTD.
XX
Roberts GW, Heal JR;
XX
WPI; 2001-408419/43.
XX
A set of peptide ligands consisting of specific complementary peptides to
proteins encoded by genes of the human genome, useful in an assay for
screening and identifying of one or more novel peptides which are drug
candidates or pro-drugs.
XX
Example 4; Page 561; 646pp; English.
XX
The invention relates to a set of complementary peptide ligands generated
from the human genome. The complementary peptides interact with their
relevant target proteins encoded in the human genome. They can be used as
reagents in drug discovery and as lead ligands to facilitate drug design
and development. The present sequence is a complementary peptide provided
in the specification
XX
SQ Sequence 10 AA;

Query Match      100.0%; Score 27; DB 4; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 RGRGR 5
Db      |||||
        5 RGRGR 9

RESULT 22
AAG97184
ID AAG97184 standard; peptide; 10 AA.
XX
AC AAG97184;
XX
DT 18-SEP-2001 (first entry)
XX
DE Human complementary peptide, SEQ ID NO: 3378.
XX
Human; complementary peptide; ligand; drug discovery; drug design.
XX
Homo sapiens.
XX
WO200142277-A2.
XX
14-JUN-2001.
XX
13-DEC-2000; 2000WO-GB004776.
XX
13-DEC-1999; 99GB-00029464.
XX
(PROT-) PROTEOM LTD.
XX
Roberts GW, Heal JR;
XX
WPI; 2001-408419/43.
XX
```

XX A set of peptide ligands consisting of specific complementary peptides to  
PT proteins encoded by genes of the human genome, useful in an assay for  
PT screening and identifying of one or more novel peptides which are drug  
PT candidates or pro-drugs.  
XX  
XX  
PS Example 4; Page 528; 646pp; English.  
XX  
XX The invention relates to a set of complementary peptide ligands generated  
CC from the human genome. The complementary peptides interact with their  
CC relevant target proteins encoded in the human genome. They can be used as  
CC reagents in drug discovery and as lead ligands to facilitate drug design  
CC and development. The present sequence is a complementary peptide provided  
CC in the specification  
XX  
XX Sequence 10 AA;  
SQ

Query Match 100.0%; Score 27; DB 4; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RGRGR 5  
DB 6 RGRGR 10

RESULT 23  
AAG95388  
ID AAG95388 standard; peptide; 10 AA.  
XX  
AC AAG95388;  
XX  
DT 18-SEP-2001 (first entry)  
XX  
DE Human complementary peptide, SEQ ID NO: 1582.  
XX  
KW Human; complementary peptide; ligand; drug discovery; drug design.  
XX  
OS Homo sapiens.  
XX  
XX WO200142277-A2.  
XX  
PD 14-JUN-2001.  
XX  
PF 13-DEC-2000; 2000WO-GB004776.  
XX  
PR 13-DEC-1999; 99GB-00029464.  
XX  
PA (PROT-) PROTEOM LTD.  
XX  
PI Roberts GW, Heal JR;  
XX  
DR WPI; 2001-408419/43.  
XX  
PT A set of peptide ligands consisting of specific complementary peptides to  
PT proteins encoded by genes of the human genome, useful in an assay for  
PT screening and identifying of one or more novel peptides which are drug  
PT candidates or pro-drugs.  
XX  
PS Example 4; Page 269; 646pp; English.  
XX  
XX The invention relates to a set of complementary peptide ligands generated  
CC from the human genome. The complementary peptides interact with their  
CC relevant target proteins encoded in the human genome. They can be used as  
CC reagents in drug discovery and as lead ligands to facilitate drug design  
CC and development. The present sequence is a complementary peptide provided  
CC in the specification  
XX  
XX Sequence 10 AA;  
SQ

Query Match 100.0%; Score 27; DB 4; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RGRGR 5  
DB 6 RGRGR 10

QY 1 RGRGR 5  
DB 1 RGRGR 5

RESULT 24  
AAG97174  
ID AAG97174 standard; peptide; 10 AA.  
XX  
AC AAG97174;  
XX  
DT 18-SEP-2001 (first entry)  
XX  
DE Human complementary peptide, SEQ ID NO: 3368.  
XX  
KW Human; complementary peptide; ligand; drug discovery; drug design.  
XX  
OS Homo sapiens.  
XX  
PN WO200142277-A2.  
XX  
PD 14-JUN-2001.  
XX  
PF 13-DEC-2000; 2000WO-GB004776.  
XX  
PR 13-DEC-1999; 99GB-00029464.  
XX  
PA (PROT-) PROTEOM LTD.  
XX  
PI Roberts GW, Heal JR;  
XX  
DR WPI; 2001-408419/43.  
XX  
PT A set of peptide ligands consisting of specific complementary peptides to  
PT proteins encoded by genes of the human genome, useful in an assay for  
PT screening and identifying of one or more novel peptides which are drug  
PT candidates or pro-drugs.  
XX  
PS Example 4; Page 527; 646pp; English.  
XX  
XX The invention relates to a set of complementary peptide ligands generated  
CC from the human genome. The complementary peptides interact with their  
CC relevant target proteins encoded in the human genome. They can be used as  
CC reagents in drug discovery and as lead ligands to facilitate drug design  
CC and development. The present sequence is a complementary peptide provided  
CC in the specification  
XX  
SQ Sequence 10 AA;  
XX

Query Match 100.0%; Score 27; DB 4; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RGRGR 5  
DB 5 RGRGR 9

RESULT 25  
AAG94474  
ID AAG94474 standard; peptide; 10 AA.  
XX  
AC AAG94474;  
XX  
DT 18-SEP-2001 (first entry)  
XX  
DE Human complementary peptide, SEQ ID NO: 668.  
XX  
KW Human; complementary peptide; ligand; drug discovery; drug design.  
XX  
OS Homo sapiens.  
XX

```
PN WO200142277-A2.
XX
XX 14-JUN-2001.
XX
XX 13-DEC-2000; 2000WO-GB004776.
XX
XX 13-DEC-1999; 99GB-00029464.
XX
XX (PROT-) PROTEOM LTD.
XX
XX Roberts GW, Heal JR;
XX
XX WPI; 2001-408419/43.
XX
XX A set of peptide ligands consisting of specific complementary peptides to
XX proteins encoded by genes of the human genome, useful in an assay for
XX screening and identifying of one or more novel peptides which are drug
XX candidates or pro-drugs.
XX
XX Example 4; Page 135; 646pp; English.
XX
XX The invention relates to a set of complementary peptide ligands generated
XX from the human genome. The complementary peptides interact with their
XX relevant target proteins encoded in the human genome. They can be used as
XX reagents in drug discovery and as lead ligands to facilitate drug design
XX and development. The present sequence is a complementary peptide provided
XX in the specification
XX
XX Sequence 10 AA;
XX
XX Query Match 100.0%; Score 27; DB 4; Length 10;
XX Best Local Similarity 100.0%; Pred. No. 1.9e+02;
XX Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 RGRGR 5
XX |||||
XX Db 5 RGRGR 9
XX
XX RESULT 26
XX AAG94478
XX ID AAG94478 standard; peptide; 10 AA.
XX
XX AC AAG94478;
XX
XX DT 18-SEP-2001 (first entry)
XX
XX DE Human complementary peptide, SEQ ID NO: 672.
XX
XX KW Human; complementary peptide; ligand; drug discovery; drug design.
XX
XX OS Homo sapiens.
XX
XX PN WO200142277-A2.
XX
XX PD 14-JUN-2001.
XX
XX PF 13-DEC-2000; 2000WO-GB004776.
XX
XX PR 13-DEC-1999; 99GB-00029464.
XX
XX PA (PROT-) PROTEOM LTD.
XX
XX PI Roberts GW, Heal JR;
XX
XX DT 18-SEP-2001 (first entry)
XX
XX DE Human complementary peptide, SEQ ID NO: 672.
XX
XX KW Human; complementary peptide; ligand; drug discovery; drug design.
XX
XX OS Homo sapiens.
XX
XX PN WO200142277-A2.
XX
XX PD 14-JUN-2001.
XX
XX PF 13-DEC-2000; 2000WO-GB004776.
XX
XX PR 13-DEC-1999; 99GB-00029464.
XX
XX PA (PROT-) PROTEOM LTD.
XX
XX PI Roberts GW, Heal JR;
XX
XX DR WPI; 2001-408419/43.
XX
XX A set of peptide ligands consisting of specific complementary peptides to
XX proteins encoded by genes of the human genome, useful in an assay for
XX screening and identifying of one or more novel peptides which are drug
XX candidates or pro-drugs.
XX
XX Example 4; Page 136; 646pp; English.
XX
XX Query Match 100.0%; Score 27; DB 4; Length 10;
XX Best Local Similarity 100.0%; Pred. No. 1.9e+02;
XX Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 RGRGR 5
XX |||||
XX Db 5 RGRGR 9
XX
XX RESULT 27
XX AAG95174
XX ID AAG95174 standard; peptide; 10 AA.
XX
XX AC AAG95174;
XX
XX DT 18-SEP-2001 (first entry)
XX
XX DE Human complementary peptide, SEQ ID NO: 1368.
XX
XX KW Human; complementary peptide; ligand; drug discovery; drug design.
XX
XX OS Homo sapiens.
XX
XX PN WO200142277-A2.
XX
XX PD 14-JUN-2001.
XX
XX PF 13-DEC-2000; 2000WO-GB004776.
XX
XX PR 13-DEC-1999; 99GB-00029464.
XX
XX PA (PROT-) PROTEOM LTD.
XX
XX PI Roberts GW, Heal JR;
XX
XX DR WPI; 2001-408419/43.
XX
XX A set of peptide ligands consisting of specific complementary peptides to
XX proteins encoded by genes of the human genome, useful in an assay for
XX screening and identifying of one or more novel peptides which are drug
XX candidates or pro-drugs.
XX
XX Example 4; Page 236; 646pp; English.
XX
XX The invention relates to a set of complementary peptide ligands generated
XX from the human genome. The complementary peptides interact with their
XX relevant target proteins encoded in the human genome. They can be used as
XX reagents in drug discovery and as lead ligands to facilitate drug design
XX and development. The present sequence is a complementary peptide provided
XX in the specification
XX
XX Sequence 10 AA;
XX
XX Query Match 100.0%; Score 27; DB 4; Length 10;
XX Best Local Similarity 100.0%; Pred. No. 1.9e+02;
XX Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 RGRGR 5
XX |||||
XX Db 6 RGRGR 10
XX
XX RESULT 28
```

AAG95528  
ID AAG95528 standard; peptide; 10 AA.  
XX  
AC AAG95528;  
XX  
DT 18-SEP-2001 (first entry)  
XX  
DE Human complementary peptide, SEQ ID NO: 1722.  
XX  
KW Human; complementary peptide; ligand; drug discovery; drug design.  
XX  
OS Homo sapiens.  
XX  
PN WO200142277-A2.  
XX  
PD 14-JUN-2001.  
XX  
PF 13-DEC-2000; 2000WO-GB004776.  
XX  
PR 13-DEC-1999; 99GB-00029464.  
XX  
PA (PROT-) PROTEOM LTD.  
XX  
PI Roberts GW, Heal JR;  
XX  
DR WPI; 2001-408419/43.  
XX  
PT A set of peptide ligands consisting of specific complementary peptides to  
PT proteins encoded by genes of the human genome, useful in an assay for  
PT screening and identifying of one or more novel peptides which are drug  
PT candidates or pro-drugs.  
XX  
PS Example 4; Page 289; 646pp; English.  
XX  
CC The invention relates to a set of complementary peptide ligands generated  
CC from the human genome. The complementary peptides interact with their  
CC relevant target proteins encoded in the human genome. They can be used as  
CC reagents in drug discovery and as lead ligands to facilitate drug design  
CC and development. The present sequence is a complementary peptide provided  
CC in the specification  
XX  
SQ Sequence 10 AA;  
Query Match 100.0%; Score 27; DB 4; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RGRGR 5  
Db |||||  
6 RGRGR 7  
RESULT 29  
AAG97180  
ID AAG97180 standard; peptide; 10 AA.  
XX  
AC AAG97180;  
XX  
DT 18-SEP-2001 (first entry)  
XX  
DE Human complementary peptide, SEQ ID NO: 3374.  
XX  
KW Human; complementary peptide; ligand; drug discovery; drug design.  
XX  
OS Homo sapiens.  
XX  
PN WO200142277-A2.  
XX  
PD 14-JUN-2001.  
XX  
PF 13-DEC-2000; 2000WO-GB004776.  
XX  
PR 13-DEC-1999; 99GB-00029464.

XX  
PA (PROT-) PROTEOM LTD.  
XX  
PI Roberts GW, Heal JR;  
XX  
DR WPI; 2001-408419/43.  
XX  
PT A set of peptide ligands consisting of specific complementary peptides to  
PT proteins encoded by genes of the human genome, useful in an assay for  
PT screening and identifying of one or more novel peptides which are drug  
PT candidates or pro-drugs.  
XX  
PS Example 4; Page 528; 646pp; English.  
XX  
CC The invention relates to a set of complementary peptide ligands generated  
CC from the human genome. The complementary peptides interact with their  
CC relevant target proteins encoded in the human genome. They can be used as  
CC reagents in drug discovery and as lead ligands to facilitate drug design  
CC and development. The present sequence is a complementary peptide provided  
CC in the specification  
XX  
SQ Sequence 10 AA;  
Query Match 100.0%; Score 27; DB 4; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RGRGR 5  
Db |||||  
6 RGRGR 10  
RESULT 30  
AAG94834  
ID AAG94834 standard; peptide; 10 AA.  
XX  
AC AAG94834;  
XX  
DT 18-SEP-2001 (first entry)  
XX  
DE Human complementary peptide, SEQ ID NO: 1028.  
XX  
KW Human; complementary peptide; ligand; drug discovery; drug design.  
XX  
OS Homo sapiens.  
XX  
PN WO200142277-A2.  
XX  
PD 14-JUN-2001.  
XX  
PF 13-DEC-2000; 2000WO-GB004776.  
XX  
PR 13-DEC-1999; 99GB-00029464.  
XX  
PA (PROT-) PROTEOM LTD.  
XX  
PI Roberts GW, Heal JR;  
XX  
DR WPI; 2001-408419/43.  
XX  
PT A set of peptide ligands consisting of specific complementary peptides to  
PT proteins encoded by genes of the human genome, useful in an assay for  
PT screening and identifying of one or more novel peptides which are drug  
PT candidates or pro-drugs.  
XX  
PS Example 4; Page 187; 646pp; English.  
XX  
CC The invention relates to a set of complementary peptide ligands generated  
CC from the human genome. The complementary peptides interact with their  
CC relevant target proteins encoded in the human genome. They can be used as  
CC reagents in drug discovery and as lead ligands to facilitate drug design  
CC and development. The present sequence is a complementary peptide provided  
CC in the specification



```
XX SQ Sequence 10 AA;
Query Match 100.0%; Score 27; DB 4; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
DB |||||
1 RGRGR 5

RESULT 31
AAG95878
ID AAG95878 standard; peptide; 10 AA.
XX
AC AAG95878;
XX
DT 18-SEP-2001 (first entry)
XX
DE Human complementary peptide, SEQ ID NO: 2072.
XX
KW Human; complementary peptide; ligand; drug discovery; drug design.
XX
OS Homo sapiens.
XX
PN WO200142277-A2.
XX
PD 14-JUN-2001.
XX
PF 13-DEC-2000; 2000WO-GB004776.
XX
PR 13-DEC-1999; 99GB-00029464.
XX
PA (PROT-) PROTEOM LTD.
XX
PI Roberts GW, Heal JR;
XX
WPI; 2001-408419/43.
XX
PT A set of peptide ligands consisting of specific complementary peptides to
PT proteins encoded by genes of the human genome, useful in an assay for
PT screening and identifying of one or more novel peptides which are drug
PT candidates or pro-drugs.
XX
PS Example 4; Page 386; 646pp; English.
XX
CC The invention relates to a set of complementary peptide ligands generated
CC from the human genome. The complementary peptides interact with their
CC relevant target proteins encoded in the human genome. They can be used as
CC reagents in drug discovery and as lead ligands to facilitate drug design
CC and development. The present sequence is a complementary peptide provided
CC in the specification
XX
SQ Sequence 10 AA;
Query Match 100.0%; Score 27; DB 4; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
DB |||||
5 RGRGR 9

RESULT 33
ADJ66128
ID ADJ66128 standard; peptide; 10 AA.
XX
AC ADJ66128;
XX
DT 06-MAY-2004 (first entry)
XX
DE Epichloe kibiensis E18 based peptide SEQ ID NO:11.
XX
KW antibacterial; biodegradable; metal-binding; drug; agrochemical;
KW cosmetic; environment-remediation.
XX
OS Synthetic.
XX
PN WO2004014944-A1.
XX
PD 19-FEB-2004.
XX
PF 18-APR-2003; 2003WO-JP004960.
XX
PR 07-AUG-2002; 2002JP-00230016.
XX
PA (OKAY-) OKAYAMA PREFECTURE.
XX
PI Nishikawa M, Ogawa K;
XX
WPI; 2004-238732/22.
XX
```

```
XX SQ Sequence 10 AA;
Query Match 100.0%; Score 27; DB 4; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
DB |||||
1 RGRGR 5

RESULT 31
AAG95878
ID AAG95878 standard; peptide; 10 AA.
XX
AC AAG95878;
XX
DT 18-SEP-2001 (first entry)
XX
DE Human complementary peptide, SEQ ID NO: 2072.
XX
KW Human; complementary peptide; ligand; drug discovery; drug design.
XX
OS Homo sapiens.
XX
PN WO200142277-A2.
XX
PD 14-JUN-2001.
XX
PF 13-DEC-2000; 2000WO-GB004776.
XX
PR 13-DEC-1999; 99GB-00029464.
XX
PA (PROT-) PROTEOM LTD.
XX
PI Roberts GW, Heal JR;
XX
WPI; 2001-408419/43.
XX
PT A set of peptide ligands consisting of specific complementary peptides to
PT proteins encoded by genes of the human genome, useful in an assay for
PT screening and identifying of one or more novel peptides which are drug
PT candidates or pro-drugs.
XX
PS Example 4; Page 339; 646pp; English.
XX
CC The invention relates to a set of complementary peptide ligands generated
CC from the human genome. The complementary peptides interact with their
CC relevant target proteins encoded in the human genome. They can be used as
CC reagents in drug discovery and as lead ligands to facilitate drug design
CC and development. The present sequence is a complementary peptide provided
CC in the specification
XX
SQ Sequence 10 AA;
Query Match 100.0%; Score 27; DB 4; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
DB |||||
3 RGRGR 7

RESULT 32
AAG96208
ID AAG96208 standard; peptide; 10 AA.
XX
AC AAG96208;
XX
DT 18-SEP-2001 (first entry)
XX
```

XX Microbe-originated polyamino-acids or their derivatives e.g.  
PT biodegradable functional polymers with antibacterial activity and metal-  
PT binding capability, useful in drugs, agrochemicals and cosmetics.  
XX  
XX Example 5; SEQ ID NO 11; 64pp; Japanese.  
XX  
XX The invention relates to novel polypeptides or their derivatives. A  
CC peptide of the invention has antibacterial activity. The polypeptides or  
CC their derivatives are particularly biodegradable functional polymers,  
CC e.g. with antibacterial activity and metal-binding capability which are  
CC useful in drugs, agrochemicals, cosmetics and environment-remediation  
CC agents. The present sequence represents a peptide of the invention.  
XX  
SQ Sequence 10 AA;  
Query Match 100.0%; Score 27; DB 8; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RGRGR 5  
Db |||||  
1 RGRGR 5  
RESULT 34  
AAW09224  
ID AAW09224 standard; peptide; 11 AA.  
XX  
AC AAW09224;  
XX  
DT 30-MAR-1997 (first entry)  
XX  
DE Repeat peptide used as polyfunctional agent for protein crosslinking.  
XX  
KW Repeat sequence; polyfunctional agent; crosslinking; protein polymer;  
KW silk fibroin; elastin; adhesive; tissue sealant; biocompatible film.  
XX  
OS Synthetic.  
XX  
PN WO9634618-A1.  
XX  
PD 07-NOV-1996.  
XX  
PF 02-MAY-1996; 96WO-US006229.  
XX  
PR 05-MAY-1995; 95US-00435641.  
XX  
PA (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.  
XX  
PI Stedronsky ER, Cappello J;  
XX  
DR WPI; 1996-505895/50.  
XX  
PT Tissue repair adhesive comprising polymer of structural protein repeat  
PT units - contg. hetero-atom functional gps. reactive with crosslinking  
PT agent, combines biocompatibility and high bonding strength.  
XX  
PS Example 4; Page 55; 103pp; English.  
XX  
CC This peptide is a repeat sequence used as a polyfunctional agent in  
CC crosslinking of new protein polymers based on silk fibroin and elastin  
CC repeat units. The synthetic peptide has been used in a composition with  
CC one of the new polymers and potassium carbonate, isocyanate and amine  
CC reagents to form tissue adhesives or sealants. In a rat skin assay, the  
CC synthetic peptide gives intermediate results when compared with other  
CC polyfunctional agents (including lysine and arginine). The new polymers  
CC may be used to seal defects in vessel walls, e.g. artery, vein,  
CC capillary, lung, dura or colon, to increase tissue mass, or to produce  
CC biocompatible films for in vivo use. The sealants have the  
CC biocompatibility of fibrin glues, but set more quickly, and have greater  
CC strength. They are readily prepared by recombinant methods, are easy to  
CC administer, and are gradually resorbed

XX Sequence 11 AA;  
SQ Query Match 100.0%; Score 27; DB 2; Length 11;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RGRGR 5  
Db |||||  
1 RGRGR 5  
RESULT 35  
AAW01039  
ID AAW01039 standard; peptide; 11 AA.  
XX  
AC AAW01039;  
XX  
DT 04-JUN-1999 (first entry)  
XX  
DE Methylated SMD homologous peptide #2.  
XX  
KW SMD; methylated SMD homologous peptide; systemic lupus erythematosus;  
KW mononucleosis; infection; cancer; Epstein-Barr virus; therapy;  
KW autoimmune disorder.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Modified-site 1..11  
FT /note= "all R are Ngamma-mono- or Ngamma-Ngamma-  
FT dimethylated arginine"  
XX  
PN WO9911667-A1.  
XX  
PD 11-MAR-1999.  
XX  
PF 31-AUG-1998; 98WO-EP005518.  
XX  
PR 29-AUG-1997; 97EP-00870127.  
XX  
PA (INNO-) INNOGENETICS NV.  
XX  
PI Meheus L, Luehrmann RG, Union A, Raymackers J;  
XX  
DR WPI; 1999-205134/17.  
XX  
PT New methylated SMD homologous peptides - useful for diagnosing and  
PT treating auto-immune diseases, and cancers associated with the Epstein-  
PT Barr virus.  
XX  
PS Claim 2; Page 42; 56pp; English.  
XX  
CC This sequence represents a methylate SMD homologous peptide of the  
CC invention. The peptides comprise an XG dimer (X = NG-mono- or NG-NG-  
CC dimethylated arginine), which react with antibodies. The antibodies,  
CC which react with the peptides, are present in sera of patients with  
CC systemic lupus erythematosus; infectious, recurrent or chronic  
CC mononucleosis or infection; or cancers associated with Epstein-Barr  
CC virus, including Burkitt's lymphoma or nasopharyngeal carcinoma. The  
CC peptides, antibodies and immunotoxin molecules form pharmaceutical  
CC compositions for diagnosing and treating autoimmune diseases (especially  
CC systemic lupus erythematosus, discoid lupus erythematosus, scleroderma,  
CC dermatomyositis, rheumatoid arthritis, Sjogren's syndrome) or diseases  
CC associated with the Epstein-Barr virus (especially Burkitt's lymphoma,  
CC nasopharyngeal carcinoma, or infectious, recurrent or chronic  
CC mononucleosis). The peptides or compositions can treat the auto-immune  
CC disease by increasing the size of antigen-immune complexes, which  
CC improves the clearance of the formed immune complexes. Methylation of the  
CC peptides permits high reactivity with antibodies present in patient sera  
CC with systemic lupus erythematosus  
XX  
SQ Sequence 11 AA;

```

Query Match      100.0%; Score 27; DB 2; Length 11;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
   |||||
Db 2 RGRGR 6

RESULT 36
AAY51894
ID AAY51894 standard; protein; 11 AA.
XX
XX AAY51894;
XX
DT 22-JUN-2000 (first entry)
XX
XX Protocol X peptide.
DE
XX Crosslinked protein; fibrin glue; tissue adhesive; sealant.
KW
XX Synthetic.
OS
XX US6033654-A.
PN
XX
XX 07-MAR-2000.
PD
XX
XX 02-MAY-1996; 96US-00642246.
PF
XX
XX 05-MAY-1995; 95US-00435641.
PR
XX
XX (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.
PA
XX
XX Cappello J, Stedronsky ER;
PI
XX
XX WPI; 2000-255682/22.
DR
XX
XX Crosslinked protein composition, useful as tissue adhesive or sealant,
PT comprises peptide repeating units that contain functional groups reactive
PT with crosslinker.
XX
XX Example 4; Col 85-86; 45pp; English.
PS
XX
XX This invention describes a novel crosslinked protein composition (A) in
CC which, before crosslinking, the protein (I) is new and comprises at least
CC 70 wt. % of repeating units GAGAGS (1) and GVGVP (2), and in at least two
CC repeating units an amino acid (aa) is substituted by Lys or Arg to
CC provide a Lys/Arg equivalent weight of 1-20 KD. (I) contains at least two
CC aa having a functional group reactive with at least one of aldehyde,
CC iso(thio)cyanate and activated carboxy. (I) have similar biocompatibility
CC to fibrin glues, but set more quickly and give a bond with greater shear
CC strength. They are made from readily available natural sources, are easy
CC to administer and are gradually resorbed. This sequence represents a
CC peptide used in protocol X of the invention
XX
XX Sequence 11 AA;
SQ
Query Match      100.0%; Score 27; DB 3; Length 11;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
   |||||
Db 1 RGRGR 5

RESULT 37
AAY90997
ID AAY90997 standard; peptide; 11 AA.
XX
XX AAY90997;
AC
XX

```

---

```

DT 05-SEP-2000 (first entry)
XX
DE Inhibitory clone p+/-4 peptide sequence SEQ ID NO:105.
XX
KW Escherichia coli; E. coli; randomised peptide library; identification;
KW stabilised bioactive peptide; synthesis; intracellular selection;
KW screening; lac operon; protease resistant; peptidase resistant;
KW Rop protein; glutathione sulphotransferase; thiodoxin; infection;
KW maltose binding protein; glutathione reductase; antimicrobial;
KW antibacterial.
XX
XX Escherichia coli.
OS Synthetic.
XX
XX WO200022112-A1.
PN
XX
XX 20-APR-2000.
PD
XX
XX 12-OCT-1999; 99WO-US023731.
PF
XX
XX 13-OCT-1998; 98US-0104013P.
PR
XX 14-DEC-1998; 98US-0112150P.
XX
XX (UYGE-) UNIV GEORGIA RES FOUND INC.
PA (ALTW/) ALTMAN E.
XX
XX Altman E;
PI
XX
XX WPI; 2000-317972/27.
DR
XX
XX Identifying recombinantly an antimicrobial bioactive peptide used as a
PT therapeutic agent involves transforming a host cell with expression
PT vector with tightly regulable control region and measuring its
PT inhibition.
XX
XX Example 3; Page 77; 135pp; English.
PS
XX
XX The present invention describes a method for identifying a bioactive
CC peptide (BP) involving transforming a cell with an expression vector
CC comprising a tightly regulatable control region operably linked to a
CC nucleic acid sequence encoding a peptide (P), growing the transformed
CC cell under conditions that repress expression of (P) and then inducing
CC its expression which, if is inhibitory to host cell growth, is indicative
CC of BP expression. An antimicrobial peptide from the present invention,
CC which is stabilised, is used for treating a patient having a condition
CC inhibiting the growth of a microbe. The new antibacterial peptides are
CC useful to treat various pathogenic bacteria such as Staphylococci,
CC Streptococci and Enterococci which are the primary causes of nosocomial
CC infections. Novel inhibitor peptides identified by the method can be
CC medical treatments and therapies directed against microbial infection.
CC Also, these novel inhibitor peptides can be used, in turn, to identify
CC additional novel antibacterial peptides using a synthetic approach, and
CC can also be used to elucidate potential new drug targets. The inhibitor
CC peptide target which is inactivated is identified using reverse genetics
CC by isolating mutants that are no longer inhibited by the peptide. These
CC mutants are then mapped in order to precisely determine the protein
CC target that is inhibited. AAA56033 to AAA56106 and AAY90964 to AAY90999
CC are sequences used in the exemplification of the present invention
XX
XX Sequence 11 AA;
SQ
Query Match      100.0%; Score 27; DB 3; Length 11;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
   |||||
Db 7 RGRGR 11

RESULT 38
ABG31424

```

ID ABG31424 standard; peptide; 11 AA.  
XX  
AC ABG31424;  
XX  
DT 29-NOV-2002 (first entry)  
XX  
DE Synthetic peptide used to evaluate properties of protein polymers.  
XX  
KW Protein polymer; functional group; crosslink; sealing; filling; tissue;  
KW tissue mass; tissue bonding; resorbable bond; flexible bond; sealant;  
KW adhesive; wound healing; burn dressing; blood flow; ruptured vessel;  
KW artery; vein; structural protein; vulnery; fibroin; elastin; collagen;  
KW keratin.  
XX  
OS Synthetic.  
XX  
PN US6423333-B1.  
XX  
PD 23-JUL-2002.  
XX  
PF 29-NOV-1999; 99US-00451206.  
XX  
PR 05-MAY-1995; 95US-00435641.  
PR 02-MAY-1996; 96US-00642246.  
XX  
PA (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.  
XX  
PI Stedronsky ER, Cappello J;  
XX  
DR WPI; 2002-672937/72.  
XX  
PT Crosslinked protein composition used as sealant or adhesive for sealing  
PT or filling defect in viable tissue, as burn dressing, or in wound healing  
PT e.g. to staunch flow of fluid e.g. blood, through ruptured vessels.  
XX  
PS Example 5; Col 38; 46pp; English.  
XX  
CC The present invention relates to protein polymers having repetitive units  
CC from naturally occurring structural proteins such as fibroin, elastin,  
CC collagen and keratin. The polymers comprise a functional group which can  
CC be chemically crosslinked with appropriate crosslinkers. The protein  
CC polymer is produced by recombinant DNA technology. The protein polymer is  
CC useful for sealing or filling a defect in viable tissue, particularly for  
CC augmenting tissue mass. The protein polymer is useful in a variety of  
CC applications related to their physical, chemical and biological  
CC properties, and/or to bond together separated tissue to provide a stable,  
CC flexible or resorbable bond. The protein is particularly useful as a  
CC sealant or adhesive, in wound healing or as a burn dressing e.g. to stop  
CC or staunch the flow of fluid (e.g. blood) through ruptured vessels (e.g.  
CC arteries or veins). The present sequence represents a synthetic peptide  
CC used to evaluate the properties of protein polymers in the examples of  
CC the present invention  
XX  
SQ Sequence 11 AA;  
  
Query Match 100.0%; Score 27; DB 5; Length 11;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 RGRGR 5  
Db 1 RGRGR 5  
  
RESULT 39  
ABW01640  
ID ABW01640 standard; peptide; 11 AA.  
XX  
AC ABW01640;  
XX  
DT 12-FEB-2004 (first entry)  
XX  
DE Synthetic peptide SEOK used to prepare HMDA setting agent.

XX Crosslinked protein; sealant; depot; drug; therapy; gel; film; thread;  
KW coating; vulnery.  
XX  
OS Synthetic.  
XX  
PN US2003104589-A1.  
XX  
PD 05-JUN-2003.  
XX  
PF 05-APR-2002; 2002US-00117931.  
XX  
PR 05-MAY-1995; 95US-00435641.  
PR 02-MAY-1996; 96US-00642246.  
PR 29-NOV-1999; 99US-00451206.  
XX  
PA (PROT-) PROTEIN POLYMER TECHNOLOGIES INC.  
XX  
PI Stedronsky ER, Cappello J;  
XX  
DR WPI; 2003-829350/77.  
XX  
PT Crosslinked protein composition useful as sealing a defect in tissue, the  
PT protein prior to crosslinking comprises repetitive units of 3-15 amino  
PT acids of natural structural protein.  
XX  
PS Example 5; Page 19; 0pp; English.  
XX  
CC The present invention relates to crosslinked protein composition. The  
CC invention is useful as sealants or depots to provide for relatively  
CC uniform release of a physiologically active product e.g. drug and for  
CC the formation of articles of manufacture such as gels, films, threads,  
CC coatings. The present sequence is synthetic peptide SEOK used to prepare  
CC HMDA setting agent  
XX  
SQ Sequence 11 AA;  
  
Query Match 100.0%; Score 27; DB 7; Length 11;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 RGRGR 5  
Db 1 RGRGR 5  
  
RESULT 40  
ADJ96780  
ID ADJ96780 standard; peptide; 11 AA.  
XX  
AC ADJ96780;  
XX  
DT 06-MAY-2004 (first entry)  
XX  
DE Stabilised bioactive opposite charge ending peptide SeqID 105.  
XX  
KW stabilised bioactive polypeptide; intracellular screening method;  
KW antibacterial; antiviral; anticancer; diagnostic tool; virucidal;  
KW cytostatic; gene therapy; opposite charge ending motif.  
XX  
OS Synthetic.  
XX  
PN WO2004011485-A2.  
XX  
PD 05-FEB-2004.  
XX  
PF 30-JUL-2003; 2003WO-US023875.  
XX  
PR 31-JUL-2002; 2002US-00210023.  
XX  
PA (UYGE-) UNIV GEORGIA RES FOUND INC.  
XX  
PI Altman E;

```

XX WPI; 2004-143816/14.
DR N-PSDB; ADJ96781.
XX
XX New stabilized bioactive peptides such as insulin, glucagon, calcitonin
PT or somatostatin, useful as therapeutic agents (e.g. as antibacterial,
PT antiviral or anticancer agents) or as diagnostic tools in research.
XX
XX Example 3; SEQ ID NO 105; 147pp; English.
XX
XX This invention relates to novel stabilised bioactive polypeptides and
CC methods of identification thereof. Specifically, it provides an
CC intracellular screening method for identifying bioactive peptides that
CC possess a stabilising group coupled to one, or both ends of the peptide,
CC such that it lacks the capacity to form an intramolecular disulphide
CC bond. The present invention describes a transformed host cell that on
CC induction can express the peptide, which as an inhibitory peptide can
CC completely or partially inhibit host cell growth (the phenotypic change
CC indicative of peptide bioactivity). As such, these bioactive peptides are
CC useful for the development of new antibacterial, antiviral or anticancer
CC agents, or as diagnostic tools in both basic and applied research.
CC Accordingly, they exhibit antibacterial, virucidal and cytostatic
CC activities and can be used for gene therapy purposes. This peptide
CC sequence is a stabilised bioactive peptide of the invention.
XX
SQ Sequence 11 AA;
Query Match 100.0%; Score 27; DB 8; Length 11;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 RGRGR 5
DB 7 RGRGR 11
|||||
|||||

RESULT 41
ADR45935
ID ADR45935 standard; peptide; 11 AA.
XX
AC ADR45935;
XX
XX 21-OCT-2004 (first entry)
XX
DE Human high mobility group HMGA2 AT-hook 1 protein.
XX
XX antidiabetic; ophthalmological; nephropathic; antiarthritic;
KW antiinflammatory; cytostatic; gynaecological; antipsoriatic;
KW dermatological; antiviral; vasotropic; antiarteriosclerotic; antianginal;
KW cardiant; vulnery; antiulcer; high mobility group protein; human.
XX
OS Homo sapiens.
XX
XX WO2004061456-A2.
XX
XX 22-JUL-2004.
XX
XX 05-JAN-2004; 2004WO-EP000030.
XX
XX 03-JAN-2003; 2003DE-01000023.
XX
XX 07-MAR-2003; 2003DE-01010160.
XX
XX 10-AUG-2003; 2003DE-01036642.
XX
XX 08-OCT-2003; 2003DE-01046614.
XX
XX (ALCE-) ALCEDO BIOTECH GMBH.
XX
XX Bullerdielk J;
XX
XX WPI; 2004-571355/55.
XX
XX N-PSDB; ADR45969.
XX
XX Use of nucleic acids encoding basic DNA-binding proteins, and their
PT translation or transcription products, for treating diseases associated

```

---

```

PT with e.g. angiogenesis, neovascularization or wound healing, also for
PT drug screening.
XX
XX Claim 57; SEQ ID NO 20; 161pp; German.
XX
XX The present invention relates to the use, especially in vitro, of nucleic
CC acids encoding high mobility group proteins for the following processes:
CC angiogenesis (including in a wound bed), neovascularization,
CC transmyocardial revascularization, wound healing, epithelialization and
CC healing in cases of tooth or bone implants. The sequences and their
CC encoded protein are useful for treating a very wide range of diseases:
CC (proliferative) diabetic retinopathy, diabetic nephropathy, macular
CC degeneration, arthritis, endometriosis, histiocytosis, psoriasis,
CC rosacea, eruptive or cavernous haemangioma, tumours (especially melanoma,
CC Kaposi sarcoma and basal cell carcinoma), haemorrhoids, arteriosclerosis,
CC angina pectoris, ischaemia, infarction, infertility, wounds (including
CC diabetic and other ulcers), also diseases associated with DNA damage,
CC especially xeroderma pigmentosum, 'leather skin', skin cancers and ageing
CC (including where associated with sunburn) and cardiac infarct, including
CC cosmetic treatment, especially as a sun-protection composition. Compounds
CC identified as modulators of the specified processes can also be used
CC therapeutically. The present sequence is a polypeptide of the invention.
XX
SQ Sequence 11 AA;
Query Match 100.0%; Score 27; DB 8; Length 11;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 RGRGR 5
DB 4 RGRGR 8
|||||
|||||

RESULT 42
ADR45929
ID ADR45929 standard; peptide; 11 AA.
XX
AC ADR45929;
XX
XX 21-OCT-2004 (first entry)
XX
DE Human high mobility group HMGA1a AT-hook 1 protein].
XX
XX antidiabetic; ophthalmological; nephropathic; antiarthritic;
KW antiinflammatory; cytostatic; gynaecological; antipsoriatic;
KW dermatological; antiviral; vasotropic; antiarteriosclerotic; antianginal;
KW cardiant; vulnery; antiulcer; high mobility group protein; human.
XX
OS Homo sapiens.
XX
XX WO2004061456-A2.
XX
XX 22-JUL-2004.
XX
XX 05-JAN-2004; 2004WO-EP000030.
XX
XX 03-JAN-2003; 2003DE-01000023.
XX
XX 07-MAR-2003; 2003DE-01010160.
XX
XX 10-AUG-2003; 2003DE-01036642.
XX
XX 08-OCT-2003; 2003DE-01046614.
XX
XX (ALCE-) ALCEDO BIOTECH GMBH.
XX
XX Bullerdielk J;
XX
XX WPI; 2004-571355/55.
XX
XX N-PSDB; ADR45963.
XX
XX Use of nucleic acids encoding basic DNA-binding proteins, and their
PT translation or transcription products, for treating diseases associated
PT with e.g. angiogenesis, neovascularization or wound healing, also for
PT drug screening.

```

XX PS Claim 57; SEQ ID NO 14; 161pp; German.

XX CC The present invention relates to the use, especially in vitro, of nucleic acids encoding high mobility group proteins for the following processes: angiogenesis (including in a wound bed), neovascularization, transmyocardial revascularization, wound healing, epithelialization and healing in cases of tooth or bone implants. The sequences and their encoded protein are useful for treating a very wide range of diseases: (proliferative) diabetic retinopathy, diabetic nephropathy, macular degeneration, arthritis, endometriosis, histiocytosis, psoriasis, rosacea, eruptive or cavernous haemangioma, tumours (especially melanoma, Kaposi sarcoma and basal cell carcinoma), haemorrhoids, arteriosclerosis, angina pectoris, ischaemia, infarction, infertility, wounds (including diabetic and other ulcers), also diseases associated with DNA damage, especially xeroderma pigmentosum, 'leather skin', skin cancers and ageing (including where associated with sunburn) and cardiac infarct, including cosmetic treatment, especially as a sun-protection composition. Compounds identified as modulators of the specified processes can also be used therapeutically. The present sequence is a polypeptide of the invention.

XX SQ Sequence 11 AA;

Query Match 100.0%; Score 27; DB 8; Length 11;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
Db 4 RGRGR 8

RESULT 43  
ADR45932  
ID ADR45932 standard; peptide; 11 AA.

XX AC ADR45932;

XX DT 21-OCT-2004 (first entry)

XX DE Human high mobility group HMGAlb AT-hook 1 protein.

XX KW antidiabetic; ophthalmological; nephropathic; antiarthritic;  
KW antiinflammatory; cytostatic; gynaecological; antipsoriatic;  
KW dermatological; antiviral; vasotropic; antiarteriosclerotic; antianginal;  
KW cardiant; vulnery; antiulcer; high mobility group protein; human.

XX OS Homo sapiens.

XX PN WO2004061456-A2.

XX PD 22-JUL-2004.

XX PF 05-JAN-2004; 2004WO-EP000030.

XX PR 03-JAN-2003; 2003DE-01000023.

XX PR 07-MAR-2003; 2003DE-01010160.

XX PR 10-AUG-2003; 2003DE-01036642.

XX PR 08-OCT-2003; 2003DE-01046614.

XX PA (ALCE-) ALCEDO BIOTECH GMBH.

XX PI Bullerdkiek J;

XX DR WPI; 2004-571355/55.

XX DR N-PSDB; ADR45966.

XX PT Use of nucleic acids encoding basic DNA-binding proteins, and their translation or transcription products, for treating diseases associated with e.g. angiogenesis, neovascularization or wound healing, also for drug screening.

XX PS Claim 57; SEQ ID NO 17; 161pp; German.

XX CC The present invention relates to the use, especially in vitro, of nucleic acids encoding high mobility group proteins for the following processes: angiogenesis (including in a wound bed), neovascularization, transmyocardial revascularization, wound healing, epithelialization and healing in cases of tooth or bone implants. The sequences and their encoded protein are useful for treating a very wide range of diseases: (proliferative) diabetic retinopathy, diabetic nephropathy, macular degeneration, arthritis, endometriosis, histiocytosis, psoriasis, rosacea, eruptive or cavernous haemangioma, tumours (especially melanoma, Kaposi sarcoma and basal cell carcinoma), haemorrhoids, arteriosclerosis, angina pectoris, ischaemia, infarction, infertility, wounds (including diabetic and other ulcers), also diseases associated with DNA damage, especially xeroderma pigmentosum, 'leather skin', skin cancers and ageing (including where associated with sunburn) and cardiac infarct, including cosmetic treatment, especially as a sun-protection composition. Compounds identified as modulators of the specified processes can also be used therapeutically. The present sequence is a polypeptide of the invention.

XX SQ Sequence 11 AA;

Query Match 100.0%; Score 27; DB 8; Length 11;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
Db 4 RGRGR 8

RESULT 44  
AAW65568  
ID AAW65568 standard; peptide; 12 AA.

XX AC AAW65568;

XX DT 27-AUG-2003 (revised)

XX DT 15-OCT-1998 (first entry)

XX DE Epstein-Barr virus derived peptide #19.

XX KW Epstein-Barr virus; diagnostic test; autoimmune disease; vaccine;  
KW infection; antibody; screening; genetic marker.

XX OS Synthetic.

XX OS Human herpesvirus 4.

XX PN WO9830586-A2.

XX PD 16-JUL-1998.

XX PF 13-JAN-1998; 98WO-US000342.

XX PR 13-JAN-1997; 97US-00781296.

XX PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.

XX PI Harley JB, James JA;

XX DR WPI; 1998-399062/34.

XX PT Use of Epstein-Barr virus or component(s) - for developing product(s) which can be used for preventing, diagnosing, treating or determining risk of developing autoimmune disease.

XX PS Claim 8; Page 64; 81pp; English.

XX CC The invention relates to a vaccine for alleviating or preventing autoimmune disorders induced by infection with Epstein-Barr virus (EBV). It comprises EBV or a component in a carrier for administration of the virus or viral component to alleviate or prevent the autoimmune disorder. Also claimed are: (1) a diagnostic test kit comprising: (a) reagents which can be used to detect levels of antibodies to EBV, indicators of

CC EBV infection of cells, or levels of EBV DNA or protein in a patient; (b)  
 CC control samples from individuals not at risk of developing an autoimmune  
 CC disease; and (c) a device for determining the differences in levels of a  
 CC patient and control samples to distinguish individuals at higher risk of  
 CC developing an autoimmune disease from those at lower risk of developing  
 CC an autoimmune disease; and (2) a method for screening for genetic markers  
 CC or risk factors for development of autoimmune disorders induced by  
 CC infection with EBV comprising comparing the responses of different  
 CC strains of the same species of an animal vaccinated with BBV or a  
 CC component to induce an autoimmune response in at least one of the strains  
 CC and comparing the differences in the genetics of the different strains to  
 CC identify potential genetic markers or risk factors. The methods can be  
 CC used for the prevention, diagnosis, and treatment of autoimmune diseases  
 CC having EBV as an etiological agent. The autoimmune diseases may be e.g.  
 CC systemic lupus erythematosus, Sjogren's syndrome, rheumatoid arthritis,  
 CC juvenile onset diabetes mellitus, Wegener's granulomatosis, etc. The  
 CC present sequence represents a peptide derived from Epstein-Barr virus.  
 CC Reagents are used to detect antibodies to this peptide in a specifically  
 CC claimed diagnostic test. (Updated on 27-AUG-2003 to correct OS field.)  
 XX  
 SQ Sequence 12 AA;

Query Match 100.0%; Score 27; DB 2; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
 |||||  
 Db 8 RGRGR 12

RESULT 45  
 AAY01044  
 ID AAY01044 standard; peptide; 12 AA.

AC AAY01044;  
 DT 04-JUN-1999 (first entry)

DE Methylated Smd homologous peptide #7.

KW Smd; methylated Smd homologous peptide; systemic lupus erythematosus;  
 KW mononucleosis; infection; cancer; Epstein-Barr virus; therapy;  
 KW autoimmune disorder.

OS Synthetic.

PH Key Location/Qualifiers  
 FT Modified-site 1..12

FT /notes= "all R are Ngamma-mono- or Ngamma-Ngamma -  
 dimethylated arginine"

PN WO9911667-Al.

XX 11-MAR-1999.

XX 31-AUG-1998; 98WO-EP005518.

XX 29-AUG-1997; 97EP-00870127.

XX (INNO-) INNOGENETICS NV.

PI Meheus L, Luehrmann RG, Union A, Raymackers J;

DR WPI; 1999-205134/17.

XX New methylated Smd homologous peptides - useful for diagnosing and  
 PT treating auto-immune diseases, and cancers associated with the Epstein-  
 PT Barr virus.

XX Claim 2; Page 42; 56pp; English.

XX This sequence represents a methylate Smd homologous peptide of the

CC invention. The peptides comprise an XG dimer (X = NG-mono- or NG-NG-  
 CC dimethylated arginine), which react with antibodies. The antibodies,  
 CC which react with the peptides, are present in sera of patients with  
 CC systemic lupus erythematosus; infectious, recurrent or chronic  
 CC mononucleosis or infection; or cancers associated with Epstein-Barr  
 CC virus, including Burkitt's lymphoma or nasopharyngeal carcinoma. The  
 CC peptides, antibodies and immunotoxin molecules form pharmaceutical  
 CC compositions for diagnosing and treating autoimmune diseases (especially  
 CC systemic lupus erythematosus, discoid lupus erythematosus, scleroderma,  
 CC dermatomyositis, rheumatoid arthritis, Sjogren's syndrome) or diseases  
 CC associated with the Epstein-Barr virus (especially Burkitt's lymphoma,  
 CC nasopharyngeal carcinoma, or infectious, recurrent or chronic  
 CC mononucleosis). The peptides or compositions can treat the auto-immune  
 CC disease by increasing the size of antigen-immune complexes, which  
 CC improves the clearance of the formed immune complexes. Methylation of the  
 CC peptides permits high reactivity with antibodies present in patient sera  
 CC with systemic lupus erythematosus  
 XX  
 SQ Sequence 12 AA;

Query Match 100.0%; Score 27; DB 2; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
 |||||  
 Db 4 RGRGR 8

RESULT 46  
 AAE09159  
 ID AAE09159 standard; peptide; 12 AA.

AC AAE09159;

XX 11-SEP-2003 (revised)

DT 15-NOV-2001 (first entry)

XX Epstein-Barr virus (BBV) peptide #4 used in the invention.

KW Vaccine; therapy; autoimmune disorder; Epstein-Barr virus; EBV; anaemia;  
 KW polymyositis; systemic lupus erythematosus; SLE; rheumatoid arthritis;  
 KW Sjogren's syndrome; diabetes mellitus; adenitis; multiple sclerosis;  
 KW demyelinating disease; Hashimoto's thyroiditis; autoimmune infertility;  
 KW hypoparathyroidism; primary biliary cirrhosis; ankylosing spondylitis;  
 KW inflammatory bowel disease; Addison's disease; thyroiditis; filariasis;  
 KW Graves' disease; ulcerative colitis; dermatomyositis; myasthenia gravis;  
 KW Crohn's disease; CREST syndrome; autoimmune cystitis; glomerulonephritis;  
 KW polyarteritis nodosa; hepatitis; atopic rhinitis; Goodpasture's syndrome;  
 KW sarcoidosis; rheumatic fever; anti-phospholipid syndrome; farmer's lung;  
 KW Cushing's syndrome; bird-fancier's lung; alveolitis; erythema nodosum;  
 KW pyoderma gangrenosum; fibromyalgia; Kawasaki's disease; cardiomyopathy;  
 KW Sampter's syndrome; asthma; polymyalgia rheumatica; psoriasis; arteritis;  
 KW erythroblastosis foetalis; cycitis; IgA nephropathy; Hodgkin's lymphoma;  
 KW renal cell carcinoma; eosinophilia; immunosuppressive; ophthalmological;  
 KW thyromimetic; neuroprotective; cytostatic; nephrotropic; antiallergic;  
 KW dengue; antiulcer; vasotropic; antipyretic; hepatotropic.

XX Human herpesvirus 4.

XX WO200158481-A2.

PD 16-AUG-2001.

XX 09-FEB-2001; 2001WO-US004191.

XX 09-FEB-2000; 2000US-00500904.

PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.

XX Harley JB, James JA, Kaufman KM;

XX WPI; 2001-522437/57.

XX Novel vaccine for alleviating or preventing autoimmune disorders induced  
PT Epstein-Barr virus (EBV) infection e.g. systemic lupus erythematosus,  
PT juvenile onset diabetes mellitus, comprises EBV virus or its component.  
XX  
PS Claim 8; Page 61; 114pp; English.  
XX  
CC The present invention relates to a vaccine for alleviating or preventing  
CC autoimmune disorders induced by infection with Epstein-Barr virus (EBV),  
CC comprising EBV or its component in a carrier. The vaccine is useful for  
CC preventing or alleviating autoimmune disorders induced by EBV, e.g.  
CC systemic lupus erythematosus (SLE), Sjogren's syndrome, juvenile onset  
CC diabetes mellitus, rheumatoid arthritis, Wegener's granulomatosis,  
CC inflammatory bowel disease, polymyositis, dermatomyositis, multiple  
CC endocrine failure, Schmidt's syndrome, autoimmune uveitis, Addison's  
CC disease, adrenailitis, primary biliary cirrhosis, Graves' disease,  
CC thyroiditis, Hashimoto's thyroiditis, autoimmune thyroid disease,  
CC pernicious and haemolytic anaemia, lupoid hepatitis, demyelinating  
CC disease, multiple sclerosis, subacute cutaneous lupus erythematosus,  
CC hypoparathyroidism, Dressler's syndrome, myasthenia gravis, autoimmune  
CC idiopathic thrombocytopenic purpura, autoimmune pemphigus vulgaris,  
CC pemphigus, bullous pemphigoid, dermatitis herpetiformis, alopecia areata,  
CC autoimmune cystitis, pemphigoid, scleroderma, progressive systemic  
CC sclerosis, CREST syndrome (Calcinosis, Raynaud's oesophageal dysmotility,  
CC sclerodactyly and telangiectasia), adult onset diabetes mellitus (Type II  
CC diabetes), male or female autoimmune infertility, ankylosing spondylitis,  
CC ulcerative colitis, Crohn's disease, mixed connective tissue disease,  
CC polyarteritis nodosa, systemic necrotising vasculitis,  
CC glomerulonephritis, atopic dermatitis, atopic rhinitis, Goodpasture's  
CC syndrome, Chagas' disease, sarcoidosis, rheumatic fever, asthma,  
CC recurrent abortion, anti-phospholipid syndrome, farmer's lung, erythema  
CC multiforme, postcardotomy syndrome, Cushing's syndrome, autoimmune  
CC chronic active hepatitis, bird-fancier's lung, allergic  
CC encephalomyelitis, toxic necrodermal lysis, alopecia, Alport's syndrome,  
CC allergic alveolitis, fibrosing alveolitis, interstitial lung disease,  
CC erythema nodosum, pyoderma gangrenosum, transfusion reaction, chronic  
CC fatigue syndrome, fibromyalgia, Takayasu's arteritis, Kawasaki's disease,  
CC polymyalgia rheumatica, temporal arteritis, giant cell arteritis, dengue,  
CC Sampter's syndrome (tridactylis, nasal polyyps, eosinophilia) and Behcet's  
CC disease, Caplan's syndrome, encephalomyositis, erythema elevatum et  
CC diuinum, psoriasis, erythroblastosis foetalis, Shulman's syndrome, Iga  
CC nephropathy, Fely's syndrome, fascitis with eosinophilia, filariasis,  
CC chronic cyclitis, heterochromic cyclitis, Fuch's cyclitis, Hodgkin's and  
CC non-Hodgkin's lymphoma, cardiomyopathy, Henoch-Schonlein purpura, post  
CC vaccination syndromes, renal cell carcinoma, Eaton-Lambert syndrome or  
CC relapsing polychondritis. The present sequence is a EBV peptide used in  
CC the invention. (Updated on 11-SEP-2003 to standardise OS field)  
XX  
SQ Sequence 12 AA;  
Query Match 100.0%; Score 27; DB 4; Length 12;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RGRGR 5  
DB 8 RGRGR 12  
RESULT 47  
ABU07643  
ID ABU07643 standard; peptide; 12 AA.  
XX AC ABU07643;  
XX  
DT 23-OCT-2003 (revised)  
DT 10-MAY-2003 (first entry)  
XX  
DE Epstein-Barr virus nuclear antigen peptide #16.  
KW EBV; viral; Epstein-Barr virus nuclear antigen; vaccine;  
KW autoimmune disorder; Epstein-Barr virus infection; diabetes mellitus;  
KW arthritis; multiple sclerosis; dermatitis; psoriasis; asthma; anaemia;

KW allergy.  
XX Human herpesvirus 4.  
XX US2002164355-A1.  
XX  
PD 07-NOV-2002.  
XX  
PF 24-OCT-2001; 2001US-00012756.  
XX  
PR 30-NOV-1993; 93US-00160604.  
PR 16-MAY-1996; 96US-0019053p.  
PR 13-JAN-1997; 97US-00781296.  
XX  
PA (HARL/) HARLEY J B.  
PA (JAME/) JAMES J A.  
XX  
PI Harley JB, James JA;  
XX  
DR WPI; 2003-298686/29.  
XX  
PT New vaccine preventing or alleviating autoimmune disorders induced by the  
PT Epstein-Barr virus, such as diabetes mellitus, rheumatoid arthritis,  
PT multiple sclerosis, systemic lupus erythematosus, atopic dermatitis and  
PT psoriasis.  
XX  
PS Claim 28; Page 28; 41pp; English.  
XX  
CC The invention relates to a vaccine for alleviating or preventing  
CC autoimmune disorders induced by infection with Epstein-Barr virus,  
CC comprising an Epstein-Barr virus or a component in a carrier for  
CC administration to alleviate or prevent the autoimmune disorders. The  
CC methods and compositions of the present invention are useful for  
CC diagnosing, preventing, treating and/or alleviating autoimmune disorders,  
CC such as diabetes mellitus, rheumatoid arthritis, juvenile rheumatoid  
CC arthritis, osteoarthritis, psoriatic arthritis, multiple sclerosis,  
CC encephalomyelitis, myasthenia gravis, systemic lupus erythematosus,  
CC autoimmune thyroiditis, atopic dermatitis, eczematous dermatitis,  
CC psoriasis, Sjogren's Syndrome, Crohn's disease, aphthous ulcer, iritis,  
CC conjunctivitis, keratoconjunctivitis, ulcerative colitis, asthma,  
CC allergic asthma, cutaneous lupus erythematosus, scleroderma, vaginitis,  
CC proctitis, drug eruptions, leprosy reversal reactions, erythema  
CC nodosum/leprosum, autoimmune uveitis, allergic encephalomyelitis, acute  
CC necrotising haemorrhagic encephalopathy, idiopathic bilateral progressive  
CC sensorineural hearing loss, aplastic anaemia, pure red cell anaemia,  
CC idiopathic thrombocytopenia, polychondritis, Wegener's granulomatosis,  
CC chronic active hepatitis, Stevens-Johnson syndrome, idiopathic sprue,  
CC lichen planus, Graves' disease, sarcoidosis, primary biliary cirrhosis,  
CC uveitis posterior, interstitial lung fibrosis, graft-versus-host disease,  
CC and allergy. The present sequence represents an Epstein-Barr virus  
CC nuclear antigen peptide used in the method of the invention. (Updated on  
CC 23-OCT-2003 to standardise OS field)  
XX  
SQ Sequence 12 AA;  
Query Match 100.0%; Score 27; DB 6; Length 12;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RGRGR 5  
DB 8 RGRGR 12  
RESULT 48  
ADO17059  
ID ADO17059 standard; peptide; 12 AA.  
XX AC ADO17059;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE Epstein-Barr virus nuclear antigen 1, antigenic peptide #7.



XX vaccine; autoimmune disorder; Epstein-Barr virus; EBV;  
 KW systemic lupus erythematosus; Sjogren's syndrome; rheumatoid arthritis;  
 KW juvenile onset diabetes mellitus; Wegener's granulomatosis;  
 KW inflammatory bowel disease; Epstein-Barr nuclear antigen 1.  
 OS Human herpesvirus 4.  
 XX  
 PN US2004086522-A1.  
 XX  
 PD 06-MAY-2004.  
 XX  
 XX 27-JUN-2003; 2003US-00607918.  
 PF  
 XX 30-NOV-1993; 93US-00160604.  
 PR 16-MAY-1996; 96US-0019053P.  
 PR 13-JAN-1997; 97US-00781296.  
 PR 24-OCT-2001; 2001US-00012756.  
 XX  
 XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
 PA  
 XX Harley JB, James JA;  
 PI  
 XX WPI; 2004-356164/33.  
 DR  
 XX  
 XX New vaccine for alleviating or preventing autoimmune disorders induced by  
 PT Epstein-Barr virus (EBV) infection, e.g. systemic lupus erythematosus,  
 PT comprises EBV or its component in a pharmaceutical carrier.  
 PT  
 XX Claim 8; Page 17; 30pp; English.  
 PS  
 XX The invention relates to a new vaccine for alleviating or preventing  
 CC autoimmune disorders induced by infection with Epstein-Barr virus (EBV).  
 CC The vaccine comprises EBV or its component in a pharmaceutical carrier  
 CC for administration of the virus or viral component in an amount and mode  
 CC of administration to alleviate or prevent the autoimmune disorders. The  
 CC composition and methods are useful for diagnosing, preventing or treating  
 CC autoimmune diseases caused by Epstein-Barr virus, such as systemic lupus  
 CC erythematosus, Sjogren's syndrome, rheumatoid arthritis, juvenile onset  
 CC diabetes mellitus, Wegener's granulomatosis or inflammatory bowel  
 CC disease. These may also be used in screening of therapeutics for  
 CC prevention or alleviation of autoimmune disorders induced by EBV  
 CC infection. The present sequence represents an Epstein-Barr virus nuclear  
 CC antigen 1, antigenic peptide used to make the vaccine of the invention.  
 CC  
 XX Sequence 12 AA;  
 SQ  
 Query Match 100.0%; Score 27; DB 8; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RGRGR 5  
 Db 8 RGRGR 12  
 RESULT 49  
 ABB82931  
 ID ABB82931 standard; peptide; 13 AA.  
 XX  
 AC ABB82931;  
 XX  
 XX 14-APR-2003 (first entry)  
 DT  
 XX Arginine-dimethylated synthetic peptide seq Id No. 1.  
 DE  
 XX Dimethylarginine; mass spectroscopy; monomethylamine; dimethylamine; DMA;  
 KW MMA; dimethylcarbodiimide; DMC.  
 XX  
 XX Synthetic.  
 OS  
 XX WO200294777-A2.  
 PN  
 XX  
 PD 28-NOV-2002.  
 XX  
 XX 20-MAY-2002; 2002WO-US015613.  
 PF  
 XX 18-MAY-2001; 2001US-0292075P.  
 PR  
 XX (MDS-) MDS PROTEOMICS INC.  
 PA  
 XX Brame CJ, Mcbroom LDB;  
 PI  
 XX WPI; 2003-140342/13.  
 DR  
 XX Identifying structure of dimethyl arginine for proteomics business by

PD 28-NOV-2002.  
 XX  
 PF 20-MAY-2002; 2002WO-US015613.  
 XX  
 XX 18-MAY-2001; 2001US-0292075P.  
 PR  
 XX (MDS-) MDS PROTEOMICS INC.  
 PA  
 XX Brame CJ, Mcbroom LDB;  
 PI  
 XX WPI; 2003-140342/13.  
 DR  
 XX Identifying structure of dimethyl arginine for proteomics business by  
 PT obtaining neutral loss spectra of peptide containing dimethylarginine  
 PT residues by mass spectrometry.  
 PT  
 XX Example 1; Page 29; 45pp; English.  
 PS  
 XX The invention relates to identifying the structure of dimethylarginine.  
 CC The method involves (i) obtaining a neutral loss spectrum of a peptide  
 CC containing a dimethylarginine by mass spectroscopy and (ii) determining  
 CC if the neutral loss spectrum shows one or both of neutral loss of  
 CC monomethylamine (MMA), dimethylcarbodiimide (DMC) and/or neutral loss of  
 CC dimethylamine (DMA). The neutral loss of MMA and DMC indicates the  
 CC presence of a symmetrically dimethylated arginine residue and the neutral  
 CC loss of (DMA) indicates the presence of asymmetrically dimethylated  
 CC arginine residue. Sequences ABB82931-939 represent synthetic peptides  
 CC having symmetrically and asymmetrically dimethylated arginine residues  
 CC used in an experiment to determine if mass spectrometry could be used  
 CC to distinguish the peptides that have symmetrically versus asymmetrically  
 CC dimethylated arginine residues  
 CC  
 XX Sequence 13 AA;  
 SQ  
 Query Match 100.0%; Score 27; DB 6; Length 13;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RGRGR 5  
 Db 4 RGRGR 8  
 RESULT 50  
 ABB82932  
 ID ABB82932 standard; peptide; 13 AA.  
 XX  
 AC ABB82932;  
 XX  
 XX 14-APR-2003 (first entry)  
 DT  
 XX Arginine-dimethylated synthetic peptide seq Id No. 2.  
 DE  
 XX Dimethylarginine; mass spectroscopy; monomethylamine; dimethylamine; DMA;  
 KW MMA; dimethylcarbodiimide; DMC.  
 XX  
 XX Synthetic.  
 OS  
 XX WO200294777-A2.  
 PN  
 XX  
 PD 28-NOV-2002.  
 XX  
 XX 20-MAY-2002; 2002WO-US015613.  
 PF  
 XX 18-MAY-2001; 2001US-0292075P.  
 PR  
 XX (MDS-) MDS PROTEOMICS INC.  
 PA  
 XX Brame CJ, Mcbroom LDB;  
 PI  
 XX WPI; 2003-140342/13.  
 DR  
 XX Identifying structure of dimethyl arginine for proteomics business by



CC presence of a symmetrically dimethylated arginine residue and the neural  
 CC loss of (DMA) indicates the presence of asymmetrically dimethylated  
 CC arginine residue. Sequences ABB82931-939 represent synthetic peptides  
 CC having symmetrically and asymmetrically dimethylated arginine residues  
 CC used in an experiment to determine if mass spectrometry could be used  
 CC to distinguish the peptides that have symmetrically versus asymmetrically  
 CC dimethylated arginine residues

XX SQ Sequence 14 AA;

Query Match 100.0%; Score 27; DB 6; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
 |||||  
 Db 4 RGRGR 8

RESULT 53

ABB82934  
 ID ABB82934 standard; peptide; 14 AA.

XX AC ABB82934;

XX DT 14-APR-2003 (first entry)

XX Arginie-dimethylated synthetic peptide seq Id No. 4.

XX Dimethylarginine; mass spectroscopy; monomethylamine; dimethylamine; DMA;  
 KW MMA; dimethylcarbodiimide; DMC.

XX Synthetic.

XX WO200294777-A2.

XX PD 28-NOV-2002.

XX PF 20-MAY-2002; 2002WO-US015613.

XX PR 18-MAY-2001; 2001US-0292075P.

XX PA (MDS- ) MDS PROTEOMICS INC.

XX PI Brame CU, Mcbroom LDB;

XX DR WPI; 2003-140342/13.

XX Identifying structure of dimethyl arginine for proteomics business by  
 PT obtaining neutral loss spectra of peptide containing dimethylarginine  
 PT residues by mass spectrometry.

XX Example 1; Page 29; 45pp; English.

XX The invention relates to identifying the structure of dimethylarginine.  
 CC The method involves (i) obtaining a neutral loss spectrum of a peptide  
 CC containing a dimethylarginine by mass spectroscopy and (ii) determining  
 CC if the neutral loss spectrum shows one or both of neutral loss of  
 CC monomethylamine (MMA), dimethylcarbodiimide (DMC) and/or neutral loss of  
 CC dimethylamine (DMA). The neutral loss of MMA and DMC indicates the  
 CC presence of a symmetrically dimethylated arginine residue and the neural  
 CC loss of (DMA) indicates the presence of asymmetrically dimethylated  
 CC arginine residue. Sequences ABB82931-939 represent synthetic peptides  
 CC having symmetrically and asymmetrically dimethylated arginine residues  
 CC used in an experiment to determine if mass spectrometry could be used  
 CC to distinguish the peptides that have symmetrically versus asymmetrically  
 CC dimethylated arginine residues

XX SQ Sequence 14 AA;

Query Match 100.0%; Score 27; DB 6; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
 |||||  
 Db 4 RGRGR 8

RESULT 54

ABB82935  
 ID ABB82935 standard; peptide; 14 AA.

XX AC ABB82935;

XX DT 14-APR-2003 (first entry)

XX Arginie-dimethylated synthetic peptide seq Id No. 5.

XX Dimethylarginine; mass spectroscopy; monomethylamine; dimethylamine; DMA;  
 KW MMA; dimethylcarbodiimide; DMC.

XX Synthetic.

XX WO200294777-A2.

XX PD 28-NOV-2002.

XX PF 20-MAY-2002; 2002WO-US015613.

XX PR 18-MAY-2001; 2001US-0292075P.

XX PA (MDS- ) MDS PROTEOMICS INC.

XX PI Brame CU, Mcbroom LDB;

XX DR WPI; 2003-140342/13.

XX Identifying structure of dimethyl arginine for proteomics business by  
 PT obtaining neutral loss spectra of peptide containing dimethylarginine  
 PT residues by mass spectrometry.

XX Example 1; Page 29; 45pp; English.

XX The invention relates to identifying the structure of dimethylarginine.  
 CC The method involves (i) obtaining a neutral loss spectrum of a peptide  
 CC containing a dimethylarginine by mass spectroscopy and (ii) determining  
 CC if the neutral loss spectrum shows one or both of neutral loss of  
 CC monomethylamine (MMA), dimethylcarbodiimide (DMC) and/or neutral loss of  
 CC dimethylamine (DMA). The neutral loss of MMA and DMC indicates the  
 CC presence of a symmetrically dimethylated arginine residue and the neural  
 CC loss of (DMA) indicates the presence of asymmetrically dimethylated  
 CC arginine residue. Sequences ABB82931-939 represent synthetic peptides  
 CC having symmetrically and asymmetrically dimethylated arginine residues  
 CC used in an experiment to determine if mass spectrometry could be used  
 CC to distinguish the peptides that have symmetrically versus asymmetrically  
 CC dimethylated arginine residues

XX SQ Sequence 14 AA;

Query Match 100.0%; Score 27; DB 6; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
 |||||  
 Db 4 RGRGR 8

RESULT 55.

ADC19846

ID ADC19846 standard; peptide; 14 AA.

XX AC ADC19846;

XX DT 18-DEC-2003 (first entry)

XX DE Fluorescently labelled spaced arginine transport peptide #26.  
XX KW Cellular membrane transport peptide; epithelial tissue;  
KW endothelial tissue; drugs transport; stratum corneum; antibacterial;  
KW antifungal; antiviral; antiproliferative; immunosuppressive; vitamin;  
KW analgesic; hormone.  
XX OS Synthetic.  
XX OS  
XX Key Location/Qualifiers  
FH Modified-site 1  
FT /label  
FT /note= "Xaa is fluorescently labelled epsilon-  
FT aminocaproic acid"  
FT Modified-site 14  
FT /label= OTHER  
FT /note= "Arg is covalently bound to a CONH2 group"  
XX US2003032593-A1.  
XX 13-FEB-2003.  
XX 14-FEB-2002; 2002US-00078247.  
XX 16-FEB-2001; 2001US-0269627P.  
XX (CELL-) CELLGATE INC.  
XX Wender PA, Rothbard JB, Wright L, Kreider EL, Vandeuken CJ;  
XX WPI; 2003-786846/74.  
XX Composition used for increasing transport of biologically active compound  
XX across biological membrane comprises biologically active compound and  
XX transport group.  
XX Example 1; Page 10; 33pp; English.  
XX The invention relates to a composition comprising a biologically active  
XX compound and a transport group. The transport group comprises a spaced  
XX poly-Arginine based peptide of formula given in the specification. The  
XX spaced poly-Arginine based peptide acts as a cellular membrane transport  
XX signal and effects transport of the biologically active compound across  
XX the membrane. The conjugate is also useful in therapeutic, prophylactic  
XX and diagnostic applications. The composition improves the transport of  
XX biologically active compounds across the biological membrane and into  
XX animal epithelial or endothelial tissues. The arginine residue of the  
XX conjugate provides an enhanced transport of drugs and are a part of the  
XX polypeptide that provides suitable spacing between arginine residues. The  
XX transport groups deliver an agent across the stratum corneum, which  
XX previously had been a nearly impenetrable barrier to drug delivery. The  
XX ability of the conjugate to obtain penetration of skin layers improves  
XX the efficacy of compounds such as antibacterials, antifungals, analgesics  
XX antivirals, antiproliferatives, immunosuppressives, vitamins, analgesics  
XX and hormones. The present sequence is a fluorescently labelled spaced  
XX arginine transport peptide of the invention.  
XX Sequence 14 AA;  
SQ Query Match 100.0%; Score 27; DB 7; Length 14;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RGRGR 5  
Db |||||  
2 RGRGR 6  
RESULT 56  
ADS52349  
ID ADS52349 standard; peptide; 14 AA.  
XX

AC ADS52349;  
XX 30-DEC-2004 (first entry)  
XX S33 peptide.  
XX S33; symmetrical dimethylated arginine; systemic lupus erythematosus;  
KW mixed connective tissue disease.  
XX Unidentified.  
XX OS  
XX Key Location/Qualifiers  
FH Misc-difference 5  
FT /note= "Symmetrical dimethylated arginine"  
XX WO2004087745-A1.  
XX 14-OCT-2004.  
XX 02-APR-2004; 2004WO-SE000526.  
XX 02-APR-2003; 2003SE-00000958.  
XX (PHAA ) PHARMACIA DIAGNOSTICS AB.  
XX Mahler M;  
XX WPI; 2004-729218/71.  
XX New peptides comprising symmetrical dimethylated arginine for diagnosing  
PT systemic lupus erythematosus (SLE) or for differentiating between SLE and  
PT mixed connective tissue diseases.  
XX Claim 2; Page 22; 38pp; English.  
XX The invention relates to a peptide (S33), comprising symmetrical  
CC dimethylated arginine (sdMA), that is able to react with antibodies which  
CC are present in sera from patients with systemic lupus erythematosus  
CC (SLE). The peptide or a multimer peptide comprising (S33) is useful for  
CC the in vitro diagnosis of systemic lupus erythematosus, for differential  
CC diagnosis to distinguish between SLE patients and patients with mixed  
CC connective tissue disease, or for the in vitro monitoring of the disease  
CC activity of dsDNA negative SLE patients. The present sequence represents  
CC the amino acid sequence of the S33 peptide.  
XX Sequence 14 AA;  
SQ Query Match 100.0%; Score 27; DB 8; Length 14;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RGRGR 5  
Db |||||  
3 RGRGR 7  
RESULT 57  
AAP60481  
ID AAP60481 standard; protein; 15 AA.  
XX AAP60481;  
XX 25-MAR-2003 (revised)  
DT 28-JUL-1991 (first entry)  
XX Epstein-Barr virus nuclear antigen (EBNA) pentapeptide.  
XX Epstein-Barr virus; nuclear antigen; passive immunization; diagnosis.  
XX Synthetic.  
XX Key Location/Qualifiers  
FH Region 4. .8  
FT

```

FT XX /label= EBNA determinant
PN XX
XX XX
XX XX
PD XX
XX XX
PF XX 02-AUG-1985; 85WO-US001484.
XX XX
XX XX 08-AUG-1984; 84US-00638726.
PR XX
XX XX (SCRI ) SCRIPPS CLINIC & RES FOUND.
PA XX
PI XX Vaughan JH, Carbon DA, Rhodes G, Houghten R;
XX XX
XX XX WPI; 1986-068962/10.
XX XX
XX XX New synthetic copolymer polypeptide(s) - useful for preventing and
PT diagnosing diseases involving Epstein-Barr virus.
PT XX
XX XX Disclosure; Page 25; 84pp; English.
XX XX
XX XX The peptide, p89(D), includes the EBNA pentapeptide which is an immunogen
CC for the production of an antibody in a host animal. Epstein-Barr virus-
CC and EBNA-associated diseases may be diagnosed and prevented. See also
CC AAP60473-P60483 (Updated on 25-MAR-2003 to correct PA field.)
XX XX
SQ XX Sequence 15 AA;

Query Match 100.0%; Score 27; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
Db |||||
3 RGRGR 7

RESULT 58
AAR25058
ID AAR25058 standard; protein; 15 AA.
XX
AC AAR25058;
XX
XX 25-MAR-2003 (revised)
DT 09-DEC-1992 (first entry)
XX
XX Synthetic random copolymer peptide P89.
DE
XX EBNA; antibody; IM; EBV; cytomegalovirus.
XX
XX Synthetic.
OS
XX US5122448-A.
PN
XX 16-JUN-1992.
PD
XX
XX 11-JAN-1990; 90US-00463505.
PF
XX
XX 08-AUG-1984; 84US-00638726.
PR
XX 24-MAR-1987; 87US-00029860.
PR 04-NOV-1987; 87US-00117241.
XX
XX (SCRI ) SCRIPPS CLINIC & RES FOUND.
PA
XX Vaughan JH, Carson DA, Rhodes G, Houghten R;
PI
XX WPI; 1992-226077/27.
DR
XX
XX Assay for anti-Epstein Barr virus nuclear antigen antibodies - using
PT random copolymer polypeptide contg. specified aminoacid residues, for
PT diagnosis of infectious mononucleosis and nasopharyngeal carcinoma.
XX
XX Disclosure; Page 16; 41pp; English.
PS

```

---

```

XX XX The sequences given in AAR25058-60 are examples of random copolymer
CC peptides which can be used in diagnostic methods and systems relating to
CC Epstein-Barr virus nuclear antigen (EBNA). These peptides are capable of
CC inducing the production of antibodies which immunoreact with EBNA, and of
CC immunoreacting with human antibodies induced by EBNA. The peptides can be
CC fixed to a solid matrix, giving a solid immunoreactant in a diagnostic
CC kit. The assay developed using these peptides has been found to be
CC clinically reliable in detecting infectious mononucleosis (IM) caused by
CC Epstein-Barr virus (EBV) as well as IM induced by cytomegalovirus, and
CC also in detecting nasopharyngeal carcinoma, another disease in which EBV
CC has been implicated. (Updated on 25-MAR-2003 to correct PF field.)
XX XX
SQ XX Sequence 15 AA;

Query Match 100.0%; Score 27; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
Db |||||
3 RGRGR 7

RESULT 59
ABP25029
ID ABP25029 standard; peptide; 15 AA.
XX
AC ABP25029;
XX
XX 11-SEP-2003 (revised)
DT 15-JUL-2002 (first entry)
XX
XX HIV DR 3b motif vpr peptide #1.
DE
XX HIV; HIV-1; human immunodeficiency virus; env; pol; gag; nef; vpr; vpu;
KW vif; tat; cytotoxic T lymphocyte; CTL; immune response; epitope; antigen;
KW vaccine; HIV infection; immunisation; virucide.
XX
XX Human immunodeficiency virus 1.
OS
XX WO200124810-A1.
PN
XX 12-APR-2001.
PD
XX 05-OCT-2000; 2000WO-US027766.
PF
XX 05-OCT-1999; 99US-00412863.
PR
XX (EPIM-) EPIMMUNE INC.
XX
XX Sette A, Sidney J, Southwood S, Livingston BD, Chesnut R;
PI Baker DM, Celis E, Kubo RT, Grey HM,
XX WPI; 2001-354887/37.
DR
XX Vaccine compositions comprising human immunodeficiency virus-1 (HIV-1)
PT peptide groups, useful for vaccinating against HIV-1.
XX
XX Claim 32; Page 409; 448pp; English.
XX
XX The present invention describes a composition (I) comprising a prepared
CC human immunodeficiency virus-1 (HIV-1) group comprising an amino acid
CC sequence selected from 51 defined amino acid sequences (ABP25347 to
CC ABP25397). (I) has virucide activity and can be used in vaccines. (I) may
CC be used for immunising subjects against HIV-1 infections. The use of
CC group-based vaccines has several advantages over traditional vaccines,
CC particularly when compared to the use of whole antigens in vaccine
CC compositions. There is evidence that the immune response to whole
CC antigens is directed largely toward variable regions of the antigen,
CC allowing for immune escape due to mutations. The groups for inclusion in
CC an group-based vaccine may be selected from conserved regions of viral or
CC tumour-associated antigens, which therefore reduces the likelihood of

```

CC escape mutants. Furthermore, immunosuppressive groups that may be present  
CC in whole antigens can be avoided with the use of group-based vaccines. An  
CC additional advantage of an group-based vaccine approach is the ability to  
CC combine selected groups (CTL and HTL), and further, to modify the  
CC composition of the groups, achieving, for example, enhanced  
CC immunogenicity. Accordingly, the immune response can be modulated, as  
CC appropriate, for the target disease. Similar engineering of the response  
CC is not possible with traditional approaches. ABP1501 to ABP25412  
CC represent peptide sequences used in the exemplification of the present  
CC invention. (Updated on 11-SEP-2003 to standardise OS field)

XX  
SQ Sequence 15 AA;

Query Match 100.0%; Score 27; DB 4; Length 15;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
DB 9 RGRGR 13

RESULT 60  
ABP51969  
ID ABP51969 standard; peptide; 15 AA.  
XX  
AC ABP51969;  
DT 09-OCT-2002 (first entry)  
XX  
DE Human RNA polymerase III subunit 21.45 N-terminal peptide SEQ ID NO:7.  
XX  
KW Human; RNA polymerase III subunit 21.45; tumour; diabetes; anaemia;  
KW embryonic development deformity; menstrual disorder; peptic ulcer;  
KW arrhythmia; epilepsy.  
XX  
OS Homo sapiens.  
XX  
FN CN1341722-A.  
XX  
PD 27-MAR-2002.  
XX  
PF 05-SEP-2000; 2000CN-00125001.  
XX  
PR 05-SEP-2000; 2000CN-00125001.  
XX  
PA (SHAN-) SHANGHAI BIODOOR GENE DEV CO LTD.  
XX  
PI Mao Y, Xie Y;  
XX  
WPI; 2002-501205/54.  
XX  
PT New polypeptide-human RNA polymerase III subunit 21.45 for treating  
PT embryonic development deformity, tumor, diabetes, menstrual disorder,  
PT peptic ulcer, arrhythmia, anemia and epilepsy.  
XX  
PS Example 6; Page 21 (Disclosure); 33pp; Chinese.  
XX  
CC The present invention describes human RNA polymerase III subunit 21.45  
CC (I). Also described is a method for producing (I) using DNA recombination  
CC technology. (I) and the polynucleotide encoding it can be used for  
CC treating several diseases, such as embryonic development deformity,  
CC tumour, diabetes, menstrual disorder, peptic ulcer, arrhythmia, anaemia  
CC and epilepsy. The present sequence represents the N-terminal peptide of  
CC (I), which can be used in an example from the present invention

XX  
SQ Sequence 15 AA;

Query Match 100.0%; Score 27; DB 5; Length 15;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5

DB 8 RGRGR 12

RESULT 61  
ADF89480  
ID ADF89480 standard; peptide; 15 AA.  
XX  
AC ADF89480;  
XX  
DT 26-FEB-2004 (first entry)  
XX  
DE Human E2F1-derived P47 peptide.  
XX  
KW histone-deacetylase; inhibitor; subtype; human; E2F1.  
XX  
OS Homo sapiens.  
XX  
FN JP2003221399-A.  
XX  
PD 05-AUG-2003.  
XX  
PF 19-NOV-2002; 2002JP-00335851.  
XX  
PR 22-NOV-2001; 2001JP-00358583.  
XX  
PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.  
XX  
DR WPI; 2003-883175/82.  
XX  
PT Novel peptide substrate or its salt, useful for measuring histone-  
PT deacetylase activity and screening of histone-deacetylase inhibitors.  
XX  
PS Disclosure; SEQ ID NO 55; 40pp; Japanese.  
XX  
CC The invention relates to a novel compound or its salt. The compound of  
CC the invention may be useful for measuring the activity of histone-  
CC deacetylase, screening for a histone-deacetylase inhibitor and  
CC identifying the subtypes of histone-deacetylase. The current sequence is  
CC that of the human E2F1-derived peptide of the invention.

XX  
SQ Sequence 15 AA;

Query Match 100.0%; Score 27; DB 7; Length 15;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
DB 7 RGRGR 11

RESULT 62  
ADF89481  
ID ADF89481 standard; peptide; 15 AA.  
XX  
AC ADF89481;  
XX  
DT 26-FEB-2004 (first entry)  
XX  
DE Human E2F1-derived P48 peptide.  
XX  
KW histone-deacetylase; inhibitor; subtype; human; E2F1.  
XX  
OS Homo sapiens.  
XX  
FN JP2003221399-A.  
XX  
PD 05-AUG-2003.  
XX  
PF 19-NOV-2002; 2002JP-00335851.  
XX  
PR 22-NOV-2001; 2001JP-00358583.

```

XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX
XX WPI; 2003-883175/82.
XX
XX Novel peptide substrate or its salt, useful for measuring histone-
PT deacetylase activity and screening of histone-deacetylase inhibitors.
XX
XX Disclosure; SEQ ID NO 56; 40pp; Japanese.
XX
XX The invention relates to a novel compound or its salt. The compound of
CC the invention may be useful for measuring the activity of histone-
CC deacetylase, screening for a histone-deacetylase inhibitor and
CC identifying the subtypes of histone-deacetylase. The current sequence is
CC that of the human E2F1-derived peptide of the invention.
XX
XX Sequence 15 AA;
XX
XX Query Match 100.0%; Score 27; DB 7; Length 15;
XX Best Local Similarity 100.0%; Pred. No. 2.6e+02;
XX Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 RGRGR 5
XX |||||
XX 4 RGRGR 8
XX
XX
XX RESULT 63
XX ADK48862
XX ID ADK48862 standard; peptide; 15 AA.
XX
XX AC ADK48862;
XX
XX 06-MAY-2004 (first entry)
XX
XX C-terminal acetylated human EFLF derived peptide SeqID 55.
XX
XX human; EFLF; erythroid Kruppel like factor; histone-deacetylase; HDAC;
XX fluorescence; colour developing substrate; histone deacetylase inhibitor;
XX cancer; cytostatic.
XX
XX Homo sapiens.
XX
XX JP2003221398-A.
XX
XX 05-AUG-2003.
XX
XX 19-NOV-2002; 2002JP-00335840.
XX
XX 22-NOV-2001; 2001JP-00358583.
XX
XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX
XX WPI; 2004-102668/11.
XX
XX Novel peptide substrate or its salt, useful for measuring histone
PT deacetylase activity.
XX
XX Disclosure; SEQ ID NO 55; 39pp; Japanese.
XX
XX This invention relates to a novel method for measuring the histone-
CC deacetylase (HDAC) activity of a substrate using specific radioactivity,
CC fluorescence or a colour developing substrate. Specifically, it refers to
CC compound such as peptidyl-Lys(Ac)-MCA, which comprises a protecting group
CC of a hydrogen atom or an amino terminus, a group in which deacetylation
CC is carried out by the histone deacetylase linked to an acetylated lysine
CC residue that in turn is linked to the fluorescent group MCA (7-amino 4-
CC methylcoumarin). The present invention describes a method for measuring
CC HDAC activity, as well as identifying the subtype of histone deacetylase.
CC Furthermore, it can be used to screen for histone deacetylase inhibitors,
CC such that they can be used to regulate gene expression. Accordingly,
CC these compositions can be used to treat cancer and exhibit cytostatic
CC activities. This peptide sequence is a C-terminal acetylated human
XX
XX Query Match 100.0%; Score 27; DB 8; Length 15;
XX Best Local Similarity 100.0%; Pred. No. 2.6e+02;
XX Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 RGRGR 5
XX |||||
XX 4 RGRGR 8
XX
XX
XX RESULT 64
XX ADK48863
XX ID ADK48863 standard; peptide; 15 AA.
XX
XX AC ADK48863;
XX
XX 06-MAY-2004 (first entry)
XX
XX C-terminal acetylated human EFLF derived peptide SeqID 56.
XX
XX human; EFLF; erythroid Kruppel like factor; histone-deacetylase; HDAC;
XX fluorescence; colour developing substrate; histone deacetylase inhibitor;
XX cancer; cytostatic.
XX
XX Homo sapiens.
XX
XX JP2003221398-A.
XX
XX 05-AUG-2003.
XX
XX 19-NOV-2002; 2002JP-00335840.
XX
XX 22-NOV-2001; 2001JP-00358583.
XX
XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX
XX WPI; 2004-102668/11.
XX
XX Novel peptide substrate or its salt, useful for measuring histone
PT deacetylase activity.
XX
XX Disclosure; SEQ ID NO 56; 39pp; Japanese.
XX
XX This invention relates to a novel method for measuring the histone-
CC deacetylase (HDAC) activity of a substrate using specific radioactivity,
CC fluorescence or a colour developing substrate. Specifically, it refers to
CC compound such as peptidyl-Lys(Ac)-MCA, which comprises a protecting group
CC of a hydrogen atom or an amino terminus, a group in which deacetylation
CC is carried out by the histone deacetylase linked to an acetylated lysine
CC residue that in turn is linked to the fluorescent group MCA (7-amino 4-
CC methylcoumarin). The present invention describes a method for measuring
CC HDAC activity, as well as identifying the subtype of histone deacetylase.
CC Furthermore, it can be used to screen for histone deacetylase inhibitors,
CC such that they can be used to regulate gene expression. Accordingly,
CC these compositions can be used to treat cancer and exhibit cytostatic
CC activities. This peptide sequence is a C-terminal acetylated human
XX
XX Query Match 100.0%; Score 27; DB 8; Length 15;
XX Best Local Similarity 100.0%; Pred. No. 2.6e+02;
XX Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 RGRGR 5
XX |||||
XX 4 RGRGR 8
XX
XX
XX RESULT 65

```

XX	02-APR-2004; 2004WO-SE000526.
XX	02-APR-2003; 2003SE-00000958.
XX	(PHAA ) PHARMACIA DIAGNOSTICS AB.
XX	Mahler M;
XX	WPI; 2004-729218/71.
XX	New peptides comprising symmetrical dimethylated arginine for diagnosing
XX	systemic lupus erythematosus (SLE) or for differentiating between SLE and
XX	mixed connective tissue diseases.
XX	Disclosure; Fig 1c; 38pp; English.
XX	The invention relates to a peptide (S33), comprising symmetrical
XX	dimechylated arginine (sDMA), that is able to react with antibodies which
XX	are present in sera from patients with systemic lupus erythematosus
XX	(SLE). The peptide or a multimer peptide comprising (S33) is useful for
XX	the in vitro diagnosis of systemic lupus erythematosus, for differential
XX	diagnosis to distinguish between SLE patients and patients with mixed
XX	connective tissue disease, or for the in vitro monitoring of the disease
XX	activity of dsDNA negative SLE patients. The present sequence represents
XX	the amino acid sequence of a Smd3 C-terminal extension peptide mimotope.
XX	Sequence 15 AA;
SQ	
	Query Match 100.0%; Score 27; DB 8; Length 15;
	Best Local Similarity 100.0%; Pred. No. 2.6e+02;
	Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 RGRGR 5
DB	3 RGRGR 7
RESULT 67	
ADS52330	
ID	ADS52330 standard; peptide; 15 AA.
XX	ADS52330;
XX	30-DEC-2004 (first entry)
DT	
XX	Smd1 C-terminal extension peptide #5.
DE	
XX	S33; symmetrical dimethylated arginine; systemic lupus erythematosus;
KW	mixed connective tissue disease; Smd1.
KW	
XX	Unidentified.
OS	
XX	WO2004087745-A1.
FN	
XX	14-OCT-2004.
PD	
XX	02-APR-2004; 2004WO-SE000526.
PF	
XX	02-APR-2003; 2003SE-00000958.
XX	(PHAA ) PHARMACIA DIAGNOSTICS AB.
PA	
XX	Mahler M;
PI	
XX	WPI; 2004-729218/71.
DR	
XX	New peptides comprising symmetrical dimethylated arginine for diagnosing
PT	systemic lupus erythematosus (SLE) or for differentiating between SLE and
PT	mixed connective tissue diseases.
XX	Disclosure; Fig 1a; 38pp; English.
XX	



CC The invention relates to a peptide (S33), comprising symmetrical  
 CC dimethylated arginine (sdMA), that is able to react with antibodies which  
 CC are present in sera from patients with systemic lupus erythematosus  
 CC (SLE). The peptide or a multimer peptide comprising (S33) is useful for  
 CC the in vitro diagnosis of systemic lupus erythematosus, for differential  
 CC diagnosis to distinguish between SLE patients and patients with mixed  
 CC connective tissue disease, or for the in vitro monitoring of the disease  
 CC activity of dsDNA negative SLE patients. The present sequence represents  
 CC the amino acid sequence of a Smd1 C-terminal extension peptide.  
 XX  
 XX Sequence 15 AA;

Query Match 100.0%; Score 27; DB 8; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
 DB 10 RGRGR 14

RESULT 68  
 ADS52335  
 ID ADS52335 standard; peptide; 15 AA.  
 XX  
 AC ADS52335;  
 XX  
 DT 30-DEC-2004 (first entry)  
 XX  
 DE Smd1 C-terminal extension peptide #10.  
 XX  
 KW S33; symmetrical dimethylated arginine; systemic lupus erythematosus;  
 KW mixed connective tissue disease; Smd1.  
 OS Unidentified.  
 XX  
 PN WO2004087745-A1.  
 XX  
 PD 14-OCT-2004.  
 XX  
 PF 02-APR-2004; 2004WO-SE000526.  
 XX  
 PR 02-APR-2003; 2003SE-00000958.  
 XX  
 PA (PHAA ) PHARMACIA DIAGNOSTICS AB.  
 XX  
 PI Mahler M;  
 XX  
 DR WPI; 2004-729218/71.  
 XX

PT New peptides comprising symmetrical dimethylated arginine for diagnosing  
 PT systemic lupus erythematosus (SLE) or for differentiating between SLE and  
 PT mixed connective tissue diseases.  
 XX  
 PS Disclosure; Fig 1a; 38pp; English.  
 XX  
 CC The invention relates to a peptide (S33), comprising symmetrical  
 CC dimethylated arginine (sdMA), that is able to react with antibodies which  
 CC are present in sera from patients with systemic lupus erythematosus  
 CC (SLE). The peptide or a multimer peptide comprising (S33) is useful for  
 CC the in vitro diagnosis of systemic lupus erythematosus, for differential  
 CC diagnosis to distinguish between SLE patients and patients with mixed  
 CC connective tissue disease, or for the in vitro monitoring of the disease  
 CC activity of dsDNA negative SLE patients. The present sequence represents  
 CC the amino acid sequence of a Smd1 C-terminal extension peptide.  
 XX  
 XX Sequence 15 AA;

Query Match 100.0%; Score 27; DB 8; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5

DB 2 RGRGR 6  
 RESULT 69  
 ADS52343  
 ID ADS52343 standard; peptide; 15 AA.  
 XX  
 AC ADS52343;  
 XX  
 DT 30-DEC-2004 (first entry)  
 XX  
 DE Smd3 C-terminal extension peptide #6.  
 XX  
 KW S33; symmetrical dimethylated arginine; systemic lupus erythematosus;  
 KW mixed connective tissue disease; Smd3.  
 OS Unidentified.  
 XX  
 PN WO2004087745-A1.  
 XX  
 PD 14-OCT-2004.  
 XX  
 PF 02-APR-2004; 2004WO-SE000526.  
 XX  
 PR 02-APR-2003; 2003SE-00000958.  
 XX  
 PA (PHAA ) PHARMACIA DIAGNOSTICS AB.  
 XX  
 PI Mahler M;  
 XX  
 DR WPI; 2004-729218/71.  
 XX  
 PT New peptides comprising symmetrical dimethylated arginine for diagnosing  
 PT systemic lupus erythematosus (SLE) or for differentiating between SLE and  
 PT mixed connective tissue diseases.  
 XX  
 PS Disclosure; Fig 1b; 38pp; English.  
 XX  
 CC The invention relates to a peptide (S33), comprising symmetrical  
 CC dimethylated arginine (sdMA), that is able to react with antibodies which  
 CC are present in sera from patients with systemic lupus erythematosus  
 CC (SLE). The peptide or a multimer peptide comprising (S33) is useful for  
 CC the in vitro diagnosis of systemic lupus erythematosus, for differential  
 CC diagnosis to distinguish between SLE patients and patients with mixed  
 CC connective tissue disease, or for the in vitro monitoring of the disease  
 CC activity of dsDNA negative SLE patients. The present sequence represents  
 CC the amino acid sequence of a Smd3 C-terminal extension peptide.  
 XX  
 XX Sequence 15 AA;

Query Match 100.0%; Score 27; DB 8; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
 DB 1 RGRGR 5

RESULT 70  
 ADS52337  
 ID ADS52337 standard; peptide; 15 AA.  
 XX  
 AC ADS52337;  
 XX  
 DT 30-DEC-2004 (first entry)  
 XX  
 DE Smd1 C-terminal extension peptide #12.  
 XX  
 KW S33; symmetrical dimethylated arginine; systemic lupus erythematosus;  
 KW mixed connective tissue disease; Smd1.  
 XX

OS Unidentified.  
 PN WO2004087745-A1.  
 XX  
 PD 14-OCT-2004.  
 XX  
 PF 02-APR-2004; 2004WO-SE000526.  
 XX  
 PR 02-APR-2003; 2003SE-00000958.  
 XX  
 PA (PHAA ) PHARMACIA DIAGNOSTICS AB.  
 XX  
 PI Mahler M;  
 XX  
 DR WPI; 2004-729218/71.  
 XX  
 PT New peptides comprising symmetrical dimethylated arginine for diagnosing  
 PT systemic lupus erythematosus (SLE) or for differentiating between SLE and  
 PT mixed connective tissue diseases.  
 XX  
 PS Disclosure; Fig 1a; 38pp; English.  
 XX  
 CC The invention relates to a peptide (S33), comprising symmetrical  
 CC dimethylated arginine (sdMA), that is able to react with antibodies which  
 CC are present in sera from patients with systemic lupus erythematosus  
 CC (SLE). The peptide or a multimer peptide comprising (S33) is useful for  
 CC the in vitro diagnosis of systemic lupus erythematosus, for differential  
 CC diagnosis to distinguish between SLE patients and patients with mixed  
 CC connective tissue disease, or for the in vitro monitoring of the disease  
 CC activity of dsDNA negative SLE patients. The present sequence represents  
 CC the amino acid sequence of a Smd1 C-terminal extension peptide.  
 XX  
 SQ Sequence 15 AA;  
 Query Match 100.0%; Score 27; DB 8; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RGRGR 5  
 DB 2 RGRGR 6  
 |||||  
 RESULT 71  
 ADS52338  
 ID ADS52338 standard; peptide; 15 AA.  
 XX  
 AC ADS52338;  
 XX  
 DT 30-DEC-2004 (first entry)  
 XX  
 DE Smd3 C-terminal extension peptide #1.  
 XX  
 KW S33; symmetrical dimethylated arginine; systemic lupus erythematosus;  
 KW mixed connective tissue disease; Smd3.  
 OS Unidentified.  
 XX  
 PN WO2004087745-A1.  
 XX  
 PD 14-OCT-2004.  
 XX  
 PF 02-APR-2004; 2004WO-SE000526.  
 XX  
 PR 02-APR-2003; 2003SE-00000958.  
 XX  
 PA (PHAA ) PHARMACIA DIAGNOSTICS AB.  
 XX  
 PI Mahler M;  
 XX  
 DR WPI; 2004-729218/71.  
 XX  
 PT New peptides comprising symmetrical dimethylated arginine for diagnosing

PT systemic lupus erythematosus (SLE) or for differentiating between SLE and  
 PT mixed connective tissue diseases.  
 XX  
 PS Disclosure; Fig 1b; 38pp; English.  
 XX  
 CC The invention relates to a peptide (S33), comprising symmetrical  
 CC dimethylated arginine (sdMA), that is able to react with antibodies which  
 CC are present in sera from patients with systemic lupus erythematosus  
 CC (SLE). The peptide or a multimer peptide comprising (S33) is useful for  
 CC the in vitro diagnosis of systemic lupus erythematosus, for differential  
 CC diagnosis to distinguish between SLE patients and patients with mixed  
 CC connective tissue disease, or for the in vitro monitoring of the disease  
 CC activity of dsDNA negative SLE patients. The present sequence represents  
 CC the amino acid sequence of a Smd3 C-terminal extension peptide.  
 XX  
 SQ Sequence 15 AA;  
 Query Match 100.0%; Score 27; DB 8; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RGRGR 5  
 DB 2 RGRGR 6  
 |||||  
 RESULT 72  
 ADS52341  
 ID ADS52341 standard; peptide; 15 AA.  
 XX  
 AC ADS52341;  
 XX  
 DT 30-DEC-2004 (first entry)  
 XX  
 DE Smd3 C-terminal extension peptide #4.  
 XX  
 KW S33; symmetrical dimethylated arginine; systemic lupus erythematosus;  
 KW mixed connective tissue disease; Smd3.  
 OS Unidentified.  
 XX  
 PN WO2004087745-A1.  
 XX  
 PD 14-OCT-2004.  
 XX  
 PF 02-APR-2004; 2004WO-SE000526.  
 XX  
 PR 02-APR-2003; 2003SE-00000958.  
 XX  
 PA (PHAA ) PHARMACIA DIAGNOSTICS AB.  
 XX  
 PI Mahler M;  
 XX  
 DR WPI; 2004-729218/71.  
 XX  
 PT New peptides comprising symmetrical dimethylated arginine for diagnosing  
 PT systemic lupus erythematosus (SLE) or for differentiating between SLE and  
 PT mixed connective tissue diseases.  
 XX  
 PS Disclosure; Fig 1b; 38pp; English.  
 XX  
 CC The invention relates to a peptide (S33), comprising symmetrical  
 CC dimethylated arginine (sdMA), that is able to react with antibodies which  
 CC are present in sera from patients with systemic lupus erythematosus  
 CC (SLE). The peptide or a multimer peptide comprising (S33) is useful for  
 CC the in vitro diagnosis of systemic lupus erythematosus, for differential  
 CC diagnosis to distinguish between SLE patients and patients with mixed  
 CC connective tissue disease, or for the in vitro monitoring of the disease  
 CC activity of dsDNA negative SLE patients. The present sequence represents  
 CC the amino acid sequence of a Smd3 C-terminal extension peptide.  
 XX  
 SQ Sequence 15 AA;  
 Query Match 100.0%; Score 27; DB 8; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RGRGR 5  
 DB 2 RGRGR 6  
 |||||  
 RESULT 72  
 ADS52341  
 ID ADS52341 standard; peptide; 15 AA.  
 XX  
 AC ADS52341;  
 XX  
 DT 30-DEC-2004 (first entry)  
 XX  
 DE Smd3 C-terminal extension peptide #4.  
 XX  
 KW S33; symmetrical dimethylated arginine; systemic lupus erythematosus;  
 KW mixed connective tissue disease; Smd3.  
 OS Unidentified.  
 XX  
 PN WO2004087745-A1.  
 XX  
 PD 14-OCT-2004.  
 XX  
 PF 02-APR-2004; 2004WO-SE000526.  
 XX  
 PR 02-APR-2003; 2003SE-00000958.  
 XX  
 PA (PHAA ) PHARMACIA DIAGNOSTICS AB.  
 XX  
 PI Mahler M;  
 XX  
 DR WPI; 2004-729218/71.  
 XX  
 PT New peptides comprising symmetrical dimethylated arginine for diagnosing  
 PT systemic lupus erythematosus (SLE) or for differentiating between SLE and  
 PT mixed connective tissue diseases.  
 XX  
 PS Disclosure; Fig 1b; 38pp; English.  
 XX  
 CC The invention relates to a peptide (S33), comprising symmetrical  
 CC dimethylated arginine (sdMA), that is able to react with antibodies which  
 CC are present in sera from patients with systemic lupus erythematosus  
 CC (SLE). The peptide or a multimer peptide comprising (S33) is useful for  
 CC the in vitro diagnosis of systemic lupus erythematosus, for differential  
 CC diagnosis to distinguish between SLE patients and patients with mixed  
 CC connective tissue disease, or for the in vitro monitoring of the disease  
 CC activity of dsDNA negative SLE patients. The present sequence represents  
 CC the amino acid sequence of a Smd3 C-terminal extension peptide.  
 XX  
 SQ Sequence 15 AA;

```

Query Match      100.0%; Score 27; DB 8; Length 15;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 RGRGR 5
       |||||
Db      3 RGRGR 7

RESULT 73
ADS52334
ID ADS52334 standard; peptide; 15 AA.
XX
AC ADS52334;
XX
DT 30-DEC-2004 (first entry)
XX
DE Smd1 C-terminal extension peptide #9.
XX
KW S33; symmetrical dimethylated arginine; systemic lupus erythematosus;
KW mixed connective tissue disease; Smd1.
XX
OS Unidentified.
XX
PN WO2004087745-A1.
XX
PD 14-OCT-2004.
XX
PF 02-APR-2004; 2004WO-SE000526.
XX
PR 02-APR-2003; 2003SE-00000958.
XX
PA (PHAA ) PHARMACIA DIAGNOSTICS AB.
XX
PI Mahler M;
XX
DR WPI; 2004-729218/71.
XX
PT New peptides comprising symmetrical dimethylated arginine for diagnosing
PT systemic lupus erythematosus (SLE) or for differentiating between SLE and
PT mixed connective tissue diseases.
XX
PS Disclosure; Fig 1a; 38pp; English.
XX
CC The invention relates to a peptide (S33), comprising symmetrical
CC dimethylated arginine (sdMA), that is able to react with antibodies which
CC are present in sera from patients with systemic lupus erythematosus
CC (SLE). The peptide or a multimer peptide comprising (S33) is useful for
CC the in vitro diagnosis of systemic lupus erythematosus, for differential
CC diagnosis to distinguish between SLE patients and patients with mixed
CC connective tissue disease, or for the in vitro monitoring of the disease
CC activity of dsDNA negative SLE patients. The present sequence represents
CC the amino acid sequence of a Smd1 C-terminal extension peptide.
XX
SQ      Sequence 15 AA;
Query Match      100.0%; Score 27; DB 8; Length 15;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 RGRGR 5
       |||||
Db      8 RGRGR 12

RESULT 75
ADS52342
ID ADS52342 standard; peptide; 15 AA.
XX
AC ADS52342;
XX
DT 30-DEC-2004 (first entry)
XX
DE Smd3 C-terminal extension peptide #5.
XX
KW S33; symmetrical dimethylated arginine; systemic lupus erythematosus;
KW mixed connective tissue disease; Smd3.
XX
OS Unidentified.
XX
PN WO2004087745-A1.
XX
PD 14-OCT-2004.
XX
PF 02-APR-2004; 2004WO-SE000526.
XX
PR 02-APR-2003; 2003SE-00000958.
XX
PA (PHAA ) PHARMACIA DIAGNOSTICS AB.
XX

```

PI Mahler M;  
XX  
DR WPI; 2004-729218/71.  
XX  
PT New peptides comprising symmetrical dimethylated arginine for diagnosing  
PT systemic lupus erythematosus (SLE) or for differentiating between SLE and  
PT mixed connective tissue diseases.  
XX  
PS Disclosure; Fig 1b; 38pp; English.  
XX  
CC The invention relates to a peptide (S33), comprising symmetrical  
CC dimethylated arginine (sdMA), that is able to react with antibodies which  
CC are present in sera from patients with systemic lupus erythematosus  
CC (SLE). The peptide or a multimer peptide comprising (S33) is useful for  
CC the in vitro diagnosis of systemic lupus erythematosus, for differential  
CC diagnosis to distinguish between SLE patients and patients with mixed  
CC connective tissue disease, or for the in vitro monitoring of the disease  
CC activity of dsDNA negative SLE patients. The present sequence represents  
CC the amino acid sequence of a Smd3 C-terminal extension peptide.  
XX  
SQ Sequence 15 AA;  
Query Match 100.0%; Score 27; DB 8; Length 15;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RGRGR 5  
Db |||||  
2 RGRGR 6  
Search completed: December 2, 2005, 10:02:41  
Job time : 157.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2005, 09:42:08 ; Search time 131.5 Seconds  
(without alignments)  
16.706 Million cell updates/sec

Title: SEQ-GRGR  
Perfect score: 27  
Sequence: 1 grgr 5

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 1000 summaries

Database : A\_Geneseq\_21.\*

- 1: Geneseqp1980s.\*
- 2: Geneseqp1990s.\*
- 3: Geneseqp2000s.\*
- 4: Geneseqp2001s.\*
- 5: Geneseqp2002s.\*
- 6: Geneseqp2003as.\*
- 7: Geneseqp2003bs.\*
- 8: Geneseqp2004s.\*
- 9: Geneseqp2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	100.0	6	2	AAW86180
2	27	100.0	7	4	ABP15087
3	27	100.0	7	4	ABP22159
4	27	100.0	7	7	ADG44442
5	27	100.0	7	7	ADG44571
6	27	100.0	7	7	ADG44425
7	27	100.0	7	7	ADG44543
8	27	100.0	7	7	ADG44655
9	27	100.0	7	7	ADG44515
10	27	100.0	7	7	ADG44669
11	27	100.0	7	7	ADG44459
12	27	100.0	7	7	ADG44457
13	27	100.0	7	7	ADG44641
14	27	100.0	7	7	ADG44473
15	27	100.0	7	7	ADG44487
16	27	100.0	7	7	ADG44613
17	27	100.0	7	7	ADG44501
18	27	100.0	7	7	ADG44599
19	27	100.0	7	7	ADG44585
20	27	100.0	7	7	ADG44529
21	27	100.0	7	7	ADG44627
22	27	100.0	8	2	AAW86181
23	27	100.0	8	4	ABP23953
24	27	100.0	8	5	ABG31118

25	27	100.0	8	7	ADD68832
26	27	100.0	8	7	ADD93307
27	27	100.0	8	8	ADT39818
28	27	100.0	8	8	ADS79236
29	27	100.0	8	8	ADT37348
30	27	100.0	8	8	ABY00351
31	27	100.0	9	4	ABP22161
32	27	100.0	10	4	ABG95908
33	27	100.0	10	4	ABP14217
34	27	100.0	10	7	ADG44611
35	27	100.0	10	7	ADG44569
36	27	100.0	10	7	ADG44423
37	27	100.0	10	7	ADG44555
38	27	100.0	10	7	ADG44625
39	27	100.0	10	7	ADG44541
40	27	100.0	10	7	ADG44471
41	27	100.0	10	7	ADG44583
42	27	100.0	10	7	ADG44457
43	27	100.0	10	7	ADG44667
44	27	100.0	10	7	ADG44440
45	27	100.0	10	7	ADG44485
46	27	100.0	10	7	ADG44513
47	27	100.0	10	7	ADG44527
48	27	100.0	10	7	ADG44639
49	27	100.0	10	7	ADG44653
50	27	100.0	10	7	ADG44597
51	27	100.0	10	7	ADG44499
52	27	100.0	11	2	AAW98451
53	27	100.0	11	2	AAW39793
54	27	100.0	11	2	AAW39794
55	27	100.0	11	2	AAW86177
56	27	100.0	11	5	ABP54079
57	27	100.0	11	7	ADC19823
58	27	100.0	11	7	ADG44414
59	27	100.0	12	2	AAW98453
60	27	100.0	12	2	AAW98452
61	27	100.0	12	2	AAW65568
62	27	100.0	12	5	ABE74766
63	27	100.0	12	5	ABE74766
64	27	100.0	12	6	ABU07643
65	27	100.0	12	8	ADO17059
66	27	100.0	13	2	AAW98450
67	27	100.0	13	2	AAW93300
68	27	100.0	13	4	AAW36917
69	27	100.0	13	4	AAW36910
70	27	100.0	13	9	ADZ38438
71	27	100.0	15	9	ADY82757
72	27	100.0	15	9	ADY82836
73	27	100.0	15	9	AEH00041
74	27	100.0	16	6	ABP82107
75	27	100.0	18	2	AAW27504
76	27	100.0	21	2	AAW27506
77	27	100.0	21	4	AAW19031
78	27	100.0	21	4	ABW38238
79	27	100.0	21	4	ABW38238
80	27	100.0	21	4	ABW38238
81	27	100.0	21	5	ABG41190
82	27	100.0	23	4	ABW36916
83	27	100.0	23	4	ABW36909
84	27	100.0	24	4	ABW36923
85	27	100.0	26	4	ABW36922
86	27	100.0	26	4	ABW36924
87	27	100.0	27	5	ABP62141
88	27	100.0	29	4	ABW36925
89	27	100.0	29	5	ADK36422
90	27	100.0	29	5	ADK36416
91	27	100.0	31	4	AAW19883
92	27	100.0	31	4	ABW39825
93	27	100.0	31	4	AAW33417
94	27	100.0	31	4	ABW34429
95	27	100.0	31	4	AAW73210
96	27	100.0	31	4	AAW60552
97	27	100.0	31	4	ABG54926

98	27	100.0	31	5	ABG43057	Abg43057 Human pep	171	27	100.0	70	4	ABB33035	Abb33035 Peptide #
99	27	100.0	32	4	ABX36926	Abx36926 Peptide #	172	27	100.0	70	4	AAM26497	Aam26497 Peptide #
100	27	100.0	33	2	ABY27508	Abx27508 E. coli b	173	27	100.0	70	4	ABB27864	Abb27864 Human pep
101	27	100.0	35	4	ABX36927	Abx36927 Peptide #	174	27	100.0	70	4	ABB18508	Abb18508 Protein #
102	27	100.0	37	6	ABO14194	Abol14194 Novel hum	175	27	100.0	70	4	AAM68220	Aam68220 Human bon
103	27	100.0	37	8	ADG78601	Adg78601 Human sec	176	27	100.0	70	4	AAU51353	Aau51353 Propionib
104	27	100.0	37	8	ADN60892	Adn60892 Human sec	177	27	100.0	70	4	AAU52795	Aau52795 Propionib
105	27	100.0	37	8	ABO57617	Abos7617 Human gen	178	27	100.0	70	4	AAM53834	Aam53834 Human bra
106	27	100.0	38	4	ABX36928	Abx36928 Peptide #	179	27	100.0	70	4	ABG47888	Abg47888 Human liv
107	27	100.0	44	4	ABB37311	Abb37311 Peptide #	180	27	100.0	70	5	ABG35870	Abg35870 Human pep
108	27	100.0	44	4	AM833215	Aam833215 Human imm	181	27	100.0	70	6	ABM47872	Abm47872 Propionib
109	27	100.0	45	4	AM211308	Aam21308 Peptide #	182	27	100.0	70	6	ABM49314	Abm49314 Propionib
110	27	100.0	45	4	ABM43641	Abm43641 Peptide #	183	27	100.0	72	4	AAAG75744	Aaag75744 Human col
111	27	100.0	45	4	AM37535	Aam37535 Peptide #	184	27	100.0	72	4	AAO11848	Aao11848 Human pol
112	27	100.0	45	4	AM77382	Aam77382 Human bon	185	27	100.0	72	5	ABP34219	Abp34219 Human ORF
113	27	100.0	45	5	ABG46396	Abg46396 Human pep	186	27	100.0	73	4	AAU51228	Aau51228 Propionib
114	27	100.0	48	2	AAI20701	Aay20701 Human neu	187	27	100.0	73	6	ABM47747	Abm47747 Propionib
115	27	100.0	48	2	AAI25802	Aay25802 Human sec	188	27	100.0	74	4	AAU46392	Aau46392 Propionib
116	27	100.0	48	7	ABM89109	Abm89109 Rice abio	189	27	100.0	74	6	ABM42911	Abm42911 Propionib
117	27	100.0	50	4	AAU66392	Aau66392 Propionib	190	27	100.0	75	4	AAU50009	Aau50009 Propionib
118	27	100.0	50	6	ABM62911	Abm62911 Propionib	191	27	100.0	75	6	ABM46528	Abm46528 Propionib
119	27	100.0	51	5	ABP63724	Abp63724 Human ORF	192	27	100.0	76	4	AM91567	Aam91567 Human imm
120	27	100.0	51	8	ADG22567	Adg22567 Cyanophag	193	27	100.0	76	4	ABG06899	Abg06899 Novel hum
121	27	100.0	52	2	AAI12712	Aay12712 Human 5'	194	27	100.0	77	3	AAAG05568	Aaag05568 Arabidops
122	27	100.0	52	4	AAU64204	Aau64204 Propionib	195	27	100.0	77	3	AAAG36816	Aaag36816 Arabidops
123	27	100.0	52	4	AAU40934	Aau40934 Propionib	196	27	100.0	77	3	AAAG04590	Aaag04590 Arabidops
124	27	100.0	52	6	ABM60723	Abm60723 Propionib	197	27	100.0	78	4	AAU42924	Aau42924 Propionib
125	27	100.0	52	6	ABM37453	Abm37453 Propionib	198	27	100.0	78	6	ABP78499	Abp78499 N. gonorr
126	27	100.0	53	4	ABG27737	Abg27737 Novel hum	199	27	100.0	78	6	ABM39443	Abm39443 Propionib
127	27	100.0	53	8	ADX68838	Adx68838 Plant ful	200	27	100.0	79	8	ADN99204	Adn99204 Novel hum
128	27	100.0	54	4	AAU40702	Aau40702 Propionib	201	27	100.0	80	4	AAU17616	Aau17616 Novel sig
129	27	100.0	54	6	ABM37221	Abm37221 Propionib	202	27	100.0	80	4	AAU64896	Aau64896 Propionib
130	27	100.0	55	4	AAU65956	Aau65956 Propionib	203	27	100.0	80	4	ABM17895	Abm17895 Human ner
131	27	100.0	55	4	AAU47438	Aau47438 Propionib	204	27	100.0	80	6	ABM61415	Abm61415 Propionib
132	27	100.0	55	6	ABM62475	Abm62475 Propionib	205	27	100.0	80	7	ADB94324	Adb94324 Human nov
133	27	100.0	55	6	ABM43957	Abm43957 Propionib	206	27	100.0	80	8	ADX79082	Adx79082 Plant ful
134	27	100.0	55	7	ADC15885	Adc15885 Human gen	207	27	100.0	81	4	AAO13674	Aao13674 Human pol
135	27	100.0	56	4	AAU58525	Aau58525 Propionib	208	27	100.0	81	5	AAU58055	Aau58055 Propionib
136	27	100.0	56	4	AAU64317	Aau64317 Propionib	209	27	100.0	81	5	ABP34986	Abp34986 Human ORF
137	27	100.0	56	4	AAU65548	Aau65548 Propionib	210	27	100.0	81	6	ABM54574	Abm54574 Propionib
138	27	100.0	56	4	AAU66939	Aau66939 Propionib	211	27	100.0	82	4	ABM51045	Abm51045 Propionib
139	27	100.0	56	5	ABG63221	Abg63221 Human pro	212	27	100.0	82	6	ABM57564	Abm57564 Propionib
140	27	100.0	56	6	ABM55044	Abm55044 Propionib	213	27	100.0	82	6	ABU19315	Abu19315 Protein e
141	27	100.0	56	6	ABM60836	Abm60836 Propionib	214	27	100.0	83	2	AAU04774	Aay04774 Mycobacte
142	27	100.0	56	6	ABM62067	Abm62067 Propionib	215	27	100.0	83	8	ADU16443	Adu16443 M. tuberc
143	27	100.0	56	6	ABM63458	Abm63458 Propionib	216	27	100.0	84	4	ABM03905	Abm03905 Human mus
144	27	100.0	56	8	ABG058563	Abg058563 Human gen	217	27	100.0	84	5	AAE23443	Aae23443 Human STI
145	27	100.0	57	4	AAU42708	Aau42708 Propionib	218	27	100.0	84	6	ABU13199	Abu13199 Novel hum
146	27	100.0	57	4	AAU63363	Aau63363 Propionib	219	27	100.0	84	8	ADJ29225	Adj29225 Human mus
147	27	100.0	57	6	ABM59882	Abm59882 Propionib	220	27	100.0	85	3	AAAG02283	Aag02283 Human sec
148	27	100.0	57	6	ABM39227	Abm39227 Propionib	221	27	100.0	85	4	ABG11374	Abg11374 Novel hum
149	27	100.0	58	4	AAU59858	Aau59858 Propionib	222	27	100.0	85	8	ADP55011	Adp55011 Human PRO
150	27	100.0	58	6	ABM56377	Abm56377 Propionib	223	27	100.0	86	4	AAU42632	Aau42632 Propionib
151	27	100.0	59	2	AAI21350	Aay21350 Human HUP	224	27	100.0	89	6	ABM39151	Abm39151 Propionib
152	27	100.0	59	4	AAE09087	Aae09087 Epstein-B	225	27	100.0	89	8	ADG22276	Adg22276 Cyanophag
153	27	100.0	59	8	ADU16732	Adu16732 M. tuberc	226	27	100.0	90	3	AAAB25506	Aab25506 Eucalyptu
154	27	100.0	60	4	AM87307	Aam87307 Human imm	227	27	100.0	91	4	ABG28472	Abg28472 Novel hum
155	27	100.0	60	8	ADU16742	Adu16742 M. tuberc	228	27	100.0	91	4	ABG17761	Abg17761 Novel hum
156	27	100.0	61	5	ADP02123	Adp02123 Human ORF	229	27	100.0	91	4	ABG17734	Abg17734 Novel hum
157	27	100.0	61	8	ABO58258	Abos8258 Human gen	230	27	100.0	91	7	ADB74648	Adb74648 Mycobacte
158	27	100.0	62	4	AAO06301	Aao06301 Human pol	231	27	100.0	92	3	AAAB69305	Aaab69305 HIV-1 non
159	27	100.0	63	8	ADG22504	Adg22504 Cyanophag	232	27	100.0	92	3	AAAG05567	Aag05567 Arabidops
160	27	100.0	64	4	AAU39910	Aau39910 Propionib	233	27	100.0	92	3	AAAG51739	Aag51739 Arabidops
161	27	100.0	64	6	ABM36429	Abm36429 Propionib	234	27	100.0	92	3	AAAG04589	Aag04589 Arabidops
162	27	100.0	64	8	ADG22271	Adg22271 Cyanophag	235	27	100.0	92	3	AAAG36815	Aag36815 Arabidops
163	27	100.0	65	3	AAAG54458	Aag54458 Zea may	236	27	100.0	92	3	AAAG04239	Aag04239 Arabidops
164	27	100.0	65	4	AAU22435	Aau22435 Human car	237	27	100.0	92	4	AAU60843	Aau60843 Propionib
165	27	100.0	65	7	ABE46403	Abex46403 Human car	238	27	100.0	92	6	ABM57362	Abm57362 Propionib
166	27	100.0	65	8	ADJ07821	Adj07821 Human car	239	27	100.0	92	7	ABO66160	Abos66160 Klebsiell
167	27	100.0	66	5	AAU81337	Aau81337 Novel hum	240	27	100.0	93	3	ABM16328	Abm16328 Eucalyptu
168	27	100.0	67	4	AM92594	Aam92594 Human dig	241	27	100.0	94	4	AAM82544	Aam82544 Human imm
169	27	100.0	68	8	ADX94889	Adx94889 Plant ful	242	27	100.0	94	4	ABG19735	Abg19735 Novel hum
170	27	100.0	70	4	AAM14090	Aam14090 Peptide #	243	27	100.0	94	6	AAO16593	Aao16593 Human Gly

244	27	100.0	95	4	Aau63876	Propionib	Aau63876	Propionib	317	27	100.0	122	4	AAM01838	Peptide #
245	27	100.0	95	6	ABM60395	Propionib	Abm60395	Propionib	318	27	100.0	122	4	ABG23977	Novel hum
246	27	100.0	96	8	ADY05867	Plant ful	Ady05867	Plant ful	319	27	100.0	122	6	ABM47391	Propionib
247	27	100.0	97	4	Aau53996	Propionib	Aau53996	Propionib	320	27	100.0	123	9	AEB00035	Novel den
248	27	100.0	97	6	ABM50515	Propionib	Abm50515	Propionib	321	27	100.0	123	2	AAR51053	Epsstein-B
249	27	100.0	97	8	ABM50515	Propionib	Abm50515	Propionib	322	27	100.0	123	2	AAR51053	Epsstein-B
250	27	100.0	98	4	ABG05365	Plant ful	Abg05365	Plant ful	323	27	100.0	123	2	AAR74990	Epstein-B
251	27	100.0	98	4	ABG05365	Novel hum	Abg05365	Novel hum	324	27	100.0	123	2	ADD01182	Human nuc
252	27	100.0	98	4	ABG22466	Novel hum	Abg22466	Novel hum	325	27	100.0	124	4	Aau40153	Propionib
253	27	100.0	99	3	AAG22677	Zea may	Aag22677	Zea may	326	27	100.0	124	4	ABG26488	Novel hum
254	27	100.0	99	3	AAG54627	Zea may	Aag54627	Zea may	327	27	100.0	124	6	ABM36672	Propionib
255	27	100.0	99	4	ABM87392	Human gen	Abm87392	Human gen	328	27	100.0	125	2	AY04858	Mycobacte
256	27	100.0	99	4	ABM70780	Drosophil	Abm70780	Drosophil	329	27	100.0	126	7	ABO63550	Klebsiell
257	27	100.0	99	5	ABG65338	Human aib	Abg65338	Human aib	330	27	100.0	126	8	ADX87410	Plant ful
258	27	100.0	99	6	ADA57334	Human sec	Ada57334	Human sec	331	27	100.0	128	3	AAI04854	Mycobacte
259	27	100.0	99	6	ADA41208	Human sec	Ada41208	Human sec	332	27	100.0	128	3	AAI04854	Mycobacte
260	27	100.0	99	6	ABR47980	Human sec	Abm47980	Human sec	333	27	100.0	129	2	AAW54067	Bone morp
261	27	100.0	99	8	ADL78665	Human sec	Adl78665	Human sec	334	27	100.0	129	2	AAW54067	Bone morp
262	27	100.0	99	8	ADL78665	Albumin f	Adl78665	Albumin f	335	27	100.0	129	8	ADU16734	M. tuberc
263	27	100.0	99	8	ABO58177	Human gen	Abm58177	Human gen	336	27	100.0	130	4	AU58171	Propionib
264	27	100.0	101	3	ABM840286	Human ORF	Abm840286	Human ORF	337	27	100.0	130	4	ABM54690	Propionib
265	27	100.0	101	4	AU471109	Propionib	Aau471109	Propionib	338	27	100.0	130	8	ADP30012	Human sec
266	27	100.0	101	5	ABP02427	Human ORF	Abp02427	Human ORF	339	27	100.0	130	8	ADP30012	Human sec
267	27	100.0	101	6	ABM43628	Propionib	Abm43628	Propionib	340	27	100.0	130	8	ADX90362	Plant ful
268	27	100.0	102	5	ABM907275	Human ORF	Abm907275	Human ORF	341	27	100.0	131	2	AY60550	Human nor
269	27	100.0	103	4	ABG22613	Novel hum	Abg22613	Novel hum	342	27	100.0	131	7	ABO83344	Pseudomon
270	27	100.0	103	8	ADP83416	Rice abio	Adp83416	Rice abio	343	27	100.0	132	7	ADC89060	Ribosomal
271	27	100.0	104	3	ABM41782	Human ORF	Abm41782	Human ORF	344	27	100.0	132	7	ADC89061	Ribosomal
272	27	100.0	104	7	ABM90261	Rice abio	Abm90261	Rice abio	345	27	100.0	133	7	ABO63403	Klebsiell
273	27	100.0	105	2	AAU04771	Mycobacte	Aau04771	Mycobacte	346	27	100.0	134	3	ABO78409	Pseudomon
274	27	100.0	105	8	ABO58643	Human gen	Abm58643	Human gen	347	27	100.0	134	3	AY79190	Haematopo
275	27	100.0	105	8	ADU16435	M. tuberc	Adu16435	M. tuberc	348	27	100.0	134	8	ADY98893	Lung spec
276	27	100.0	106	2	AAU04796	Mycobacte	Aau04796	Mycobacte	349	27	100.0	134	9	ABE40083	L. pneumo
277	27	100.0	106	7	ABO62886	Klebsiell	Abm62886	Klebsiell	350	27	100.0	135	7	ABO75414	Pseudomon
278	27	100.0	106	7	ABO64902	Klebsiell	Abm64902	Klebsiell	351	27	100.0	136	7	ABO75414	Pseudomon
279	27	100.0	106	8	ADU16505	M. tuberc	Adu16505	M. tuberc	352	27	100.0	136	7	ABO74472	Pseudomon
280	27	100.0	107	3	AGG01410	Human sec	Aag01410	Human sec	353	27	100.0	136	7	ABO74472	Pseudomon
281	27	100.0	107	4	ABM84240	Human imm	Abm84240	Human imm	354	27	100.0	137	4	ABM64607	Putative
282	27	100.0	108	4	AAU39688	Propionib	Aau39688	Propionib	355	27	100.0	137	6	ABM64607	Putative
283	27	100.0	108	6	ABM36207	Propionib	Abm36207	Propionib	356	27	100.0	137	7	ADC87869	Ribosomal
284	27	100.0	110	3	ABM41890	Human ORF	Abm41890	Human ORF	357	27	100.0	137	7	ADC89137	Ribosomal
285	27	100.0	110	4	AAU54131	Propionib	Aau54131	Propionib	358	27	100.0	137	7	ADC88056	Ribosomal
286	27	100.0	110	5	ABP07698	Human ORF	Abp07698	Human ORF	359	27	100.0	137	7	ADC88056	Ribosomal
287	27	100.0	110	5	ABP09415	Human ORF	Abp09415	Human ORF	360	27	100.0	137	8	ADK64528	Disease t
288	27	100.0	111	6	ABM50650	Propionib	Abm50650	Propionib	361	27	100.0	138	6	ADA54654	Human pro
289	27	100.0	111	4	AAU38984	Human pol	Aau38984	Human pol	362	27	100.0	138	6	ADA54654	Human pro
290	27	100.0	112	4	AAU23390	Novel hum	Aau23390	Novel hum	363	27	100.0	138	7	ADK61860	Disease t
291	27	100.0	112	4	AAU23390	Novel hum	Aau23390	Novel hum	364	27	100.0	138	7	ADM05291	Human pro
292	27	100.0	112	4	AAU81630	Human hae	Aau81630	Human hae	365	27	100.0	138	7	ABO77122	Pseudomon
293	27	100.0	112	4	AAU81869	Human hae	Aau81869	Human hae	366	27	100.0	138	8	ADX92584	Plant ful
294	27	100.0	112	4	AAU67621	Propionib	Aau67621	Propionib	367	27	100.0	139	4	AAU65979	Propionib
295	27	100.0	113	3	AAU09173	Propionib	Aau09173	Propionib	368	27	100.0	139	6	ABM62498	Propionib
296	27	100.0	113	3	AAU09173	Hepatit	Aau09173	Hepatit	369	27	100.0	139	7	ABO72586	Pseudomon
297	27	100.0	115	8	ADL04410	Cyanophag	Adl04410	Cyanophag	370	27	100.0	140	8	ADN47738	Thermococ
298	27	100.0	115	8	ADL04410	M. catarr	Adl04410	M. catarr	371	27	100.0	140	8	ADN47738	Thermococ
299	27	100.0	115	8	ADX90752	Plant ful	Adx90752	Plant ful	372	27	100.0	141	3	AAU6486	M. tuberc
300	27	100.0	116	7	ABM63940	Human pro	Abm63940	Human pro	373	27	100.0	141	3	AAU6486	M. tuberc
301	27	100.0	117	4	AAU42905	Propionib	Aau42905	Propionib	374	27	100.0	141	4	ABM35444	Peptide #
302	27	100.0	117	6	ABM39424	Propionib	Abm39424	Propionib	375	27	100.0	141	4	ABM35444	Peptide #
303	27	100.0	117	7	ADW26604	Hyperther	Adm26604	Hyperther	376	27	100.0	141	4	ABM35444	Peptide #
304	27	100.0	117	8	ADY05845	Plant ful	Ady05845	Plant ful	377	27	100.0	141	4	ABM35444	Peptide #
305	27	100.0	118	4	AAO03518	Human pol	Aao03518	Human pol	378	27	100.0	141	4	ABM35444	Peptide #
306	27	100.0	118	7	ADC30984	Human nov	Adc30984	Human nov	379	27	100.0	141	4	ABM35444	Peptide #
307	27	100.0	121	2	AAU12319	Human 5'	Aau12319	Human 5'	380	27	100.0	141	4	ABM35444	Peptide #
308	27	100.0	121	7	ADW03867	Human pro	Adm03867	Human pro	381	27	100.0	141	5	ABG38223	Human pep
309	27	100.0	122	2	AAW78494	C-termina	Aaw78494	C-termina	382	27	100.0	141	5	ABG38223	Human pep
310	27	100.0	122	4	AAU14100	Peptide #	Aau14100	Peptide #	383	27	100.0	141	7	ABO75052	Pseudomon
311	27	100.0	122	4	ABM33045	Peptide #	Abm33045	Peptide #	384	27	100.0	141	7	ABO75052	Pseudomon
312	27	100.0	122	4	AAU26507	Peptide #	Aau26507	Peptide #	385	27	100.0	142	4	AAU26507	Peptide #
313	27	100.0	122	4	ABM27874	Human pep	Abm27874	Human pep	386	27	100.0	142	4	AAU26507	Peptide #
314	27	100.0	122	4	AAU66230	Human bon	Aau66230	Human bon	387	27	100.0	142	5	ABP66678	Human bre
315	27	100.0	122	4	AAU50872	Propionib	Aau50872	Propionib	388	27	100.0	142	5	ABP66678	Human bre
316	27	100.0	122	4	AAU53843	Human bra	Aau53843	Human bra	389	27	100.0	142	8	ADJ09386	Human pro

390	27	100.0	143	2	AAY04787	Aay04787 Mycobacte	463	27	100.0	155	7	ABO69061	Ab069061 Pseudom
391	27	100.0	143	7	ADF28883	Adf28883 Human nor	464	27	100.0	156	8	ADP84559	Adp84559 Human bre
392	27	100.0	143	7	AB082209	Ab082209 Pseudom	465	27	100.0	156	8	ADY12807	Ady12807 Plant ful
393	27	100.0	143	7	AB072266	Ab072266 Pseudom	466	27	100.0	157	7	ABO80789	Ab080789 Pseudom
394	27	100.0	143	8	ADY05984	Ady05984 Plant ful	467	27	100.0	157	8	ADY07092	Ady07092 Plant ful
395	27	100.0	144	7	AB078192	Ab078192 Pseudom	468	27	100.0	158	3	AAG44865	Aag44865 Zea mays
396	27	100.0	144	7	AB072755	Ab072755 Pseudom	469	27	100.0	158	4	ABB15089	Abb15089 Human ner
397	27	100.0	145	4	ABB50641	Abb50641 Human sec	470	27	100.0	158	7	ADJ70903	Adj70903 Human hea
398	27	100.0	145	6	AB044898	Ab044898 Novel hum	471	27	100.0	158	7	AB075837	Ab075837 Pseudom
399	27	100.0	145	7	AB026378	Ab026378 Protein a	472	27	100.0	158	8	ADX78679	Adx78679 Plant ful
400	27	100.0	145	7	AB071286	Ab071286 Pseudom	473	27	100.0	159	8	AB083166	Ab083166 Pseudom
401	27	100.0	146	5	AAM51934	Aam51934 Murine TG	474	27	100.0	159	8	ADX72246	Adx72246 Plant ful
402	27	100.0	146	7	AB076380	Ab076380 Pseudom	475	27	100.0	160	7	AB070621	Ab070621 Pseudom
403	27	100.0	146	8	ADX69020	Adx69020 Plant ful	476	27	100.0	160	8	ADG22381	Adg22381 Cyanophag
404	27	100.0	147	4	AAW95469	Aaw95469 Human rep	477	27	100.0	160	8	ADY05748	Ady05748 Plant ful
405	27	100.0	147	4	ABB96154	Abb96154 Human tes	478	27	100.0	160	8	ADX90488	Adx90488 Plant ful
406	27	100.0	147	7	AB084075	Ab084075 Pseudom	479	27	100.0	161	2	AAR65182	Aar65182 GDF-7 C-t
407	27	100.0	148	7	AB074997	Ab074997 Pseudom	480	27	100.0	161	3	AAG44856	Aag44856 Zea mays
408	27	100.0	149	7	ABW73680	Abw73680 DNA clone	481	27	100.0	161	7	ABO68966	Ab068966 Pseudom
409	27	100.0	149	7	ADC88053	Adc88053 Ribosomal	482	27	100.0	161	7	ABO81235	Ab081235 Pseudom
410	27	100.0	149	7	AB071498	Ab071498 Pseudom	483	27	100.0	161	8	ADX90753	Adx90753 Plant ful
411	27	100.0	150	3	ABR25212	AbR25212 Eucalyptu	484	27	100.0	162	7	ABO68249	Ab068249 Pseudom
412	27	100.0	150	3	AAG04588	Aag04588 Arabidops	485	27	100.0	162	8	ADG22583	Adg22583 Cyanophag
413	27	100.0	150	3	AAG51738	Aag51738 Arabidops	486	27	100.0	163	2	AAW70458	Aaw70458 Human TI-
414	27	100.0	150	3	AAG05566	Aag05566 Arabidops	487	27	100.0	163	8	ADP29391	Adp29391 Human sec
415	27	100.0	150	3	AAG36814	Aag36814 Arabidops	488	27	100.0	163	8	ADX94051	Adx94051 Plant ful
416	27	100.0	150	3	AAG04238	Aag04238 Arabidops	489	27	100.0	164	4	ABG04352	Abg04352 Novel hum
417	27	100.0	150	4	AAO05710	Aao05710 Human pol	490	27	100.0	164	7	ABO67984	Ab067984 Pseudom
418	27	100.0	150	7	ADC88055	Adc88055 Ribosomal	491	27	100.0	165	4	ABG22474	Abg22474 Novel hum
419	27	100.0	150	7	ADC89138	Adc89138 Ribosomal	492	27	100.0	165	7	ABO68855	Ab068855 Pseudom
420	27	100.0	150	7	AB076399	Ab076399 Pseudom	493	27	100.0	165	7	ABO67235	Ab067235 Klebsiell
421	27	100.0	150	7	AB079053	Ab079053 Pseudom	494	27	100.0	166	4	AAG78621	Aag78621 Human S18
422	27	100.0	151	8	ADN74313	Adn74313 Thale cre	495	27	100.0	166	4	ABB70468	Abb70468 Drosophil
423	27	100.0	151	3	AAG03940	Aag03940 Human sec	496	27	100.0	166	4	AAW83703	Aaw83703 Human imm
424	27	100.0	151	3	ABB58223	Abb58223 Drosophil	497	27	100.0	166	8	ADX96014	Adx96014 Plant ful
425	27	100.0	151	4	ABB58216	Abb58216 Drosophil	498	27	100.0	167	8	AAW79515	Aaw79515 Human pro
426	27	100.0	151	4	ABB65704	Abb65704 Drosophil	499	27	100.0	167	5	ABG60156	Abg60156 Human DIT
427	27	100.0	151	6	ABG76038	Abg76038 Human GDF	500	27	100.0	167	8	ADX75557	Adx75557 Plant ful
428	27	100.0	151	7	ABW74349	Abw74349 DNA clone	501	27	100.0	167	8	ADX95934	Adx95934 Plant ful
429	27	100.0	151	7	ADC89136	Adc89136 Ribosomal	502	27	100.0	167	8	ADX88899	Adx88899 Plant ful
430	27	100.0	151	7	ADC89132	Adc89132 Ribosomal	503	27	100.0	167	8	ADX66794	Adx66794 Plant ful
431	27	100.0	151	7	ADC89148	Adc89148 Ribosomal	504	27	100.0	167	8	ADY06746	Ady06746 Plant ful
432	27	100.0	151	7	ADC87930	Adc87930 Ribosomal	505	27	100.0	167	8	ADX74048	Adx74048 Plant ful
433	27	100.0	151	7	ADC89147	Adc89147 Ribosomal	506	27	100.0	167	8	ADY22956	Ady22956 Plant ful
434	27	100.0	151	7	ADC89149	Adc89149 Ribosomal	507	27	100.0	167	8	ADX67919	Adx67919 Plant ful
435	27	100.0	151	7	ADP58917	Adp58917 Human pol	508	27	100.0	167	8	ADX77626	Adx77626 Plant ful
436	27	100.0	151	7	ADK11352	Adk11352 Human cyt	509	27	100.0	168	8	ADX95678	Adx95678 Plant ful
437	27	100.0	151	7	ADK11348	Adk11348 Drosophil	510	27	100.0	168	8	ADX67104	Adx67104 Plant ful
438	27	100.0	151	7	ADK11350	Adk11350 Drosophil	511	27	100.0	168	8	ADY05608	Ady05608 Plant ful
439	27	100.0	151	7	ADJ68861	Adj68861 Human hea	512	27	100.0	168	8	ADX66801	Adx66801 Plant ful
440	27	100.0	151	7	ABO68420	Ab068420 Pseudom	513	27	100.0	170	7	ABO77560	Ab077560 Pseudom
441	27	100.0	151	8	ADG14199	Adg14199 Mouse ful	514	27	100.0	170	8	ADO62476	Ado62476 Transcrip
442	27	100.0	151	8	ADP30015	Adp30015 Human sec	515	27	100.0	171	7	ABM89518	Abm89518 Rice abio
443	27	100.0	151	8	ABM80659	Abm80659 Tumour-as	516	27	100.0	171	8	ADG22770	Adg22770 Cyanophag
444	27	100.0	151	8	ADP74329	Adp74329 PRO polyP	517	27	100.0	171	8	ADX88125	Adx88125 Plant ful
445	27	100.0	151	8	ADT60475	Adt60475 plant pol	518	27	100.0	171	8	ADX88126	Adx88126 Plant ful
446	27	100.0	151	9	ADY15982	Ady15982 PRO polyP	519	27	100.0	171	8	ADX77611	Adx77611 Plant ful
447	27	100.0	152	3	AAG26931	Aag26931 Zea mays	520	27	100.0	171	8	ADX77642	Adx77642 Plant ful
448	27	100.0	152	3	ABB64065	Abb64065 Drosophil	521	27	100.0	172	4	AAW96995	Aaw96995 Human com
449	27	100.0	152	4	AAW78531	Aaw78531 Human pro	522	27	100.0	172	7	ABO71788	Ab071788 Pseudom
450	27	100.0	152	4	ADC89229	Adc89229 Ribosomal	523	27	100.0	172	7	ABO79495	Ab079495 Pseudom
451	27	100.0	152	7	ADC89130	Adc89130 Ribosomal	524	27	100.0	172	7	ABO75735	Ab075735 Pseudom
452	27	100.0	152	7	ADC89234	Adc89234 Ribosomal	525	27	100.0	172	7	ABO69537	Ab069537 Pseudom
453	27	100.0	152	7	ADC89146	Adc89146 Ribosomal	526	27	100.0	172	8	ADY23010	Ady23010 Plant ful
454	27	100.0	152	8	ADO57239	Ado57239 Kidney de	527	27	100.0	173	4	ABG11561	Abg11561 Novel hum
455	27	100.0	152	8	ADQ65166	Adq65166 Novel hum	528	27	100.0	173	7	ABO79403	Ab079403 Pseudom
456	27	100.0	152	8	ABM80754	Abm80754 Tumour-as	529	27	100.0	173	8	ADX79390	Adx79390 Plant ful
457	27	100.0	152	9	ADX05716	Adx05716 Cyclin-de	530	27	100.0	173	8	ADY11592	Ady11592 Plant ful
458	27	100.0	153	7	ADC89131	Adc89131 Ribosomal	531	27	100.0	174	5	ABP42310	Abp42310 Human ova
459	27	100.0	153	7	ABO74732	Ab074732 Pseudom	532	27	100.0	176	7	ABO72106	Ab072106 Pseudom
460	27	100.0	154	7	ADB64713	Adb64713 Human pro	533	27	100.0	176	7	ABO80766	Ab080766 Pseudom
461	27	100.0	155	7	ADC89231	Adc89231 Ribosomal	534	27	100.0	176	8	ADX80182	Adx80182 Plant ful
462	27	100.0	155	7	ABO73282	Ab073282 Pseudom	535	27	100.0	176	9	AEA13749	Aea13749 Human alp



536	27	100.0	177	5	ABU67203	G-protein	Abu67203	609	27	100.0	195	2	AAV07105	Colon can
537	27	100.0	177	6	ABP53984	Human adr	Abp53984	610	27	100.0	196	7	ABO74900	Pseudomon
538	27	100.0	177	7	ABO69358	Pseudomon	Abp69358	611	27	100.0	196	8	ADY12774	Plant ful
539	27	100.0	177	8	ADO28724	Human alp	Ado28724	612	27	100.0	198	7	ABO78641	Pseudomon
540	27	100.0	177	8	ADO05228	Adrenergi	Ado05228	613	27	100.0	199	8	ADY24871	Plant ful
541	27	100.0	177	9	ADX44543	Human cla	Adx44543	614	27	100.0	199	8	ADX78672	Plant ful
542	27	100.0	177	9	ADY83779	Alpha-1B-	Ady83779	615	27	100.0	199	8	ADX71891	Plant ful
543	27	100.0	177	9	AEAl3816	Human alp	Aea13816	616	27	100.0	199	8	ADY25283	Plant ful
544	27	100.0	178	4	AAU42272	Propionib	Aau42272	617	27	100.0	200	6	ABP56645	Chimpanze
545	27	100.0	178	6	ABM38791	Propionib	Abm38791	618	27	100.0	200	6	ADY23159	Plant ful
546	27	100.0	178	7	ABO74967	Pseudomon	Abp74967	619	27	100.0	201	8	ADY27100	Bacterial
547	27	100.0	178	7	ABO61719	Klebsiell	Abp61719	620	27	100.0	201	8	ADY26716	Bacterial
548	27	100.0	178	9	AEA20055	Novel hum	Aea20055	621	27	100.0	201	8	ADY26348	Bacterial
549	27	100.0	179	4	AAm18625	Peptide #	Aam18625	622	27	100.0	202	8	ADY57401	Plant pol
550	27	100.0	179	4	ABB37682	Peptide #	Abb37682	623	27	100.0	204	7	ABO76310	Pseudomon
551	27	100.0	179	4	AAW31085	Peptide #	Aaw31085	624	27	100.0	204	8	ADY7687	Pseudomon
552	27	100.0	179	4	AAW70787	Human bon	Aaw70787	625	27	100.0	204	8	ADY94579	Plant ful
553	27	100.0	179	4	AAW58313	Human bra	Aaw58313	626	27	100.0	205	7	ABM88661	Rice abio
554	27	100.0	179	4	ABG52437	Human liv	Abg52437	627	27	100.0	206	7	ABO77456	Pseudomon
555	27	100.0	179	5	ABG40524	Human pep	Abg40524	628	27	100.0	206	8	ADY29261	Plant ful
556	27	100.0	179	7	ABO78158	Pseudomon	Abp78158	629	27	100.0	207	7	ABO80759	Pseudomon
557	27	100.0	180	7	ABO69840	Pseudomon	Abp69840	630	27	100.0	208	8	ADY79550	Plant ful
558	27	100.0	180	7	ABO73068	Pseudomon	Abp73068	631	27	100.0	209	3	AAG43767	Zea maye
559	27	100.0	180	8	ABO58503	Human gen	Abp58503	632	27	100.0	209	4	AAU67923	Propionib
560	27	100.0	181	7	ADJ11416	Rice prot	Adj11416	633	27	100.0	209	6	ABM64442	Propionib
561	27	100.0	181	7	ADJ11770	Rice prot	Adj11770	634	27	100.0	209	7	ABO67930	Pseudomon
562	27	100.0	181	7	ABO74339	Pseudomon	Abp74339	635	27	100.0	209	8	ABO84466	Human can
563	27	100.0	182	3	AAG51737	Arabidops	Aag51737	636	27	100.0	209	8	ADY57069	Plant pol
564	27	100.0	182	4	ABG26542	Novel hum	Abg26542	637	27	100.0	210	8	ADO65398	Novel hum
565	27	100.0	182	4	ABG20369	Novel hum	Abg20369	638	27	100.0	210	8	ADY92525	Plant ful
566	27	100.0	182	7	ABO82050	Pseudomon	Abp82050	639	27	100.0	212	7	ABO75766	Pseudomon
567	27	100.0	182	7	ABO61784	Pseudomon	Abp61784	640	27	100.0	212	8	ADN06066	Antipeori
568	27	100.0	182	8	ADY92908	Plant ful	Ady92908	641	27	100.0	212	8	ADY24956	Plant ful
569	27	100.0	182	8	ADX90948	Plant ful	Adx90948	642	27	100.0	212	8	ADY05490	Plant ful
570	27	100.0	183	1	AAP60624	Sequence	Aap60624	643	27	100.0	213	4	ABAl9897	Polyketid
571	27	100.0	183	4	ABG16356	Novel hum	Abg16356	644	27	100.0	213	7	ABO79564	Pseudomon
572	27	100.0	183	7	ADE09130	Novel pro	Ade09130	645	27	100.0	214	7	ABO82172	Pseudomon
573	27	100.0	183	7	ABO74979	Pseudomon	Abp74979	646	27	100.0	215	8	ADY72868	Plant ful
574	27	100.0	184	3	AAG04237	Arabidops	Aag04237	647	27	100.0	215	8	ADY24349	Plant ful
575	27	100.0	184	4	ABG29375	Novel hum	Abg29375	648	27	100.0	215	8	ADX88512	Plant ful
576	27	100.0	184	5	AAE15263	Human RNA	Aae15263	649	27	100.0	215	8	ADX88202	Plant ful
577	27	100.0	184	6	ABP80748	N. gonorr	Abp80748	650	27	100.0	216	5	AAU72506	Arabidops
578	27	100.0	184	6	AAO16589	Human Gly	Aao16589	651	27	100.0	216	7	ABO77915	Pseudomon
579	27	100.0	184	6	ADA33606	Acinetoba	Ada33606	652	27	100.0	217	4	AAU63038	Propionib
580	27	100.0	184	7	ABO70242	Pseudomon	Abp70242	653	27	100.0	217	6	ABP76929	N. gonorr
581	27	100.0	184	8	ADX66505	Plant ful	Adx66505	654	27	100.0	217	6	ABP79994	N. gonorr
582	27	100.0	185	7	ADC08149	Rice prot	Adc08149	655	27	100.0	217	6	ABP77925	N. gonorr
583	27	100.0	185	7	ADP75595	Pseudomon	Adp75595	656	27	100.0	217	6	ABP78744	N. gonorr
584	27	100.0	186	7	ADP59012	Human pol	Adp59012	657	27	100.0	217	6	ABP78635	N. gonorr
585	27	100.0	186	7	ABO72262	Pseudomon	Abp72262	658	27	100.0	217	6	ABP77011	N. gonorr
586	27	100.0	186	7	ABO81912	Pseudomon	Abp81912	659	27	100.0	217	6	ABP79136	N. gonorr
587	27	100.0	186	8	ADY06150	Plant ful	Ady06150	660	27	100.0	217	6	ABP78855	N. gonorr
588	27	100.0	186	8	ADY96818	Plant ful	Ady96818	661	27	100.0	217	6	ABP78954	N. gonorr
589	27	100.0	187	7	ABO75271	Pseudomon	Abp75271	662	27	100.0	217	6	ABP78700	N. gonorr
590	27	100.0	187	7	ABO74280	Pseudomon	Abp74280	663	27	100.0	217	6	ABP77913	N. gonorr
591	27	100.0	187	8	ADY05808	Plant ful	Ady05808	664	27	100.0	217	6	ABP80757	N. gonorr
592	27	100.0	188	6	AAO16596	Human Gly	Aao16596	665	27	100.0	217	6	ABP77228	N. gonorr
593	27	100.0	188	6	ADA48094	Rice prot	Ada48094	666	27	100.0	217	6	ABM59557	Propionib
594	27	100.0	189	4	ABE64383	Drosophi	Abbe64383	667	27	100.0	217	7	ABO79066	Pseudomon
595	27	100.0	190	4	AAU64837	Propionib	Aau64837	668	27	100.0	217	8	ADY08299	Plant ful
596	27	100.0	190	6	ABM61356	Propionib	Abm61356	669	27	100.0	218	8	ABO58433	Human gen
597	27	100.0	191	6	ABP77841	N. gonorr	Abp77841	670	27	100.0	218	7	ADM26625	Novel hum
598	27	100.0	191	6	ABP78785	N. gonorr	Abp78785	671	27	100.0	219	8	ADY09558	Hyperther
599	27	100.0	191	7	ABO75769	Pseudomon	Abp75769	672	27	100.0	220	4	ABG16831	Novel hum
600	27	100.0	191	7	ABO67847	Pseudomon	Abp67847	673	27	100.0	220	8	ADY96028	Plant ful
601	27	100.0	191	8	ADJ48383	Maize oil	Adj48383	674	27	100.0	220	8	ADX72381	Plant ful
602	27	100.0	191	8	ADP30013	Human sec	Adp30013	675	27	100.0	222	4	ABG29970	Novel hum
603	27	100.0	193	7	ABO68196	Pseudomon	Abp68196	676	27	100.0	222	4	ABG29970	Novel hum
604	27	100.0	193	7	ABM89155	Rice abio	Abm89155	677	27	100.0	223	8	ADY09558	Pseudomon
605	27	100.0	193	8	ADR98892	Lung sec	Adr98892	678	27	100.0	224	6	ADA48140	Rice prot
606	27	100.0	193	8	ADY12691	Plant ful	Ady12691	679	27	100.0	224	8	ADU02539	Novel hum
607	27	100.0	194	4	ABG15399	Novel hum	Abg15399	680	27	100.0	224	8	ADX71938	Plant ful
608	27	100.0	194	8	ADY05509	Plant ful	Ady05509	681	27	100.0	225	7	ABO69008	Pseudomon

682	27	100.0	227	7	ADM26236	Adm26236	Hyperther
683	27	100.0	227	7	ABO75698	AbO75698	Pseudomon
684	27	100.0	227	7	ABO74460	AbO74460	Pseudomon
685	27	100.0	227	8	ADX74121	AdX74121	Plant ful
686	27	100.0	228	7	ABO79585	AbO79585	Pseudomon
687	27	100.0	228	7	ABO78618	AbO78618	Pseudomon
688	27	100.0	229	4	AAU01359	AaU01359	Human TAN
689	27	100.0	229	4	ABG90579	AbG90579	Human sec
690	27	100.0	229	5	ABG65458	AbG65458	Human alb
691	27	100.0	229	8	ADL78725	AdL78725	Albumin f
692	27	100.0	230	7	ABO76982	AbO76982	Pseudomon
693	27	100.0	231	4	ABG12370	AbG12370	Novel hum
694	27	100.0	231	8	ADX91501	AdX91501	Plant ful
695	27	100.0	233	6	ABO53011	AbO53011	Human spl
696	27	100.0	233	7	ADF66174	AdF66174	Human par
697	27	100.0	233	8	ADQ21725	AdQ21725	Human sof
698	27	100.0	234	8	AAU113008	AaU113008	Plant ful
699	27	100.0	235	4	AAU011380	AaU011380	Human TAN
700	27	100.0	235	8	ADY23058	AdY23058	Plant ful
701	27	100.0	236	3	AG29858	Aag29858	Arabidops
702	27	100.0	236	7	ABO77234	AbO77234	Pseudomon
703	27	100.0	238	7	ADF58723	AdF58723	Human pol
704	27	100.0	238	7	ADM25747	Adm25747	Hyperther
705	27	100.0	239	7	ABO72806	AbO72806	Pseudomon
706	27	100.0	240	2	ABR78738	Aar78738	Murine mV
707	27	100.0	240	2	AAW26594	Aaw26594	Murine BM
708	27	100.0	240	4	AAE10984	Aae10984	Murine pa
709	27	100.0	240	6	ABG73300	AbG73300	Amino aci
710	27	100.0	240	8	ADY23473	AdY23473	Plant ful
711	27	100.0	240	8	ADY13143	AdY13143	Plant ful
712	27	100.0	240	8	ADX92867	AdX92867	Plant ful
713	27	100.0	241	4	AB59843	Ab59843	Drosophil
714	27	100.0	242	4	ABB70949	Abb70949	Pseudomon
715	27	100.0	242	8	ADY08343	AdY08343	Plant ful
716	27	100.0	243	7	ADB65540	AdB65540	Human pro
717	27	100.0	244	7	ABO73530	AbO73530	Pseudomon
718	27	100.0	244	8	ADX87644	AdX87644	Plant ful
719	27	100.0	244	8	ADX68219	AdX68219	Plant ful
720	27	100.0	245	3	AA73441	Aay73441	Human gen
721	27	100.0	245	4	AB87413	AbB7413	Human gen
722	27	100.0	245	4	AAU01382	AaU01382	Human TAN
723	27	100.0	245	4	AB90580	Ab90580	Human sec
724	27	100.0	245	4	AB90547	Ab90547	Human sec
725	27	100.0	245	4	AB90578	Ab90578	Human sec
726	27	100.0	245	4	AB90581	Ab90581	Human sec
727	27	100.0	245	4	AB90577	Ab90577	Human sec
728	27	100.0	245	5	ABG65399	AbG65399	Human alb
729	27	100.0	245	5	ABG65456	AbG65456	Human alb
730	27	100.0	245	5	ABG65457	AbG65457	Human alb
731	27	100.0	245	5	ABG65455	AbG65455	Human alb
732	27	100.0	245	5	ABG65454	AbG65454	Human alb
733	27	100.0	245	5	ABG65452	AbG65452	Human alb
734	27	100.0	245	6	ABU98379	Abu98379	Human com
735	27	100.0	245	6	ADA56775	Ada56775	Human sec
736	27	100.0	245	6	ADA57335	Ada57335	Human sec
737	27	100.0	245	6	ADA40623	Ada40623	Human sec
738	27	100.0	245	6	ADA41209	Ada41209	Human sec
739	27	100.0	245	6	ABR47981	AbR47981	Human sec
740	27	100.0	245	6	ABR47660	AbR47660	Human sec
741	27	100.0	245	7	ADC01980	AdC01980	Human com
742	27	100.0	245	7	ADC74036	AdC74036	Human sec
743	27	100.0	245	7	ADC74411	AdC74411	Human sec
744	27	100.0	245	8	ADL78666	AdL78666	Albumin f
745	27	100.0	245	8	ADL78723	AdL78723	Albumin f
746	27	100.0	245	8	ADL78719	AdL78719	Albumin f
747	27	100.0	245	8	ADL78722	AdL78722	Albumin f
748	27	100.0	245	8	ADL78724	AdL78724	Albumin f
749	27	100.0	245	8	ADP78721	AdP78721	Albumin f
750	27	100.0	245	8	ADP24915	AdP24915	PRO polyP
751	27	100.0	245	8	ADS10663	AdS10663	Human the
752	27	100.0	245	8	ADX74313	AdX74313	Plant ful
753	27	100.0	246	8	ADT57595	AdT57595	Plant pol
754	27	100.0	246	8	ADX87771	AdX87771	Plant ful

755	27	100.0	246	8	ADX77272	AdX77272	Plant ful
756	27	100.0	247	7	ADM04049	Adm04049	Human pro
757	27	100.0	247	8	ADT60718	AdT60718	Plant pol
758	27	100.0	248	4	AAU01383	AaU01383	Human TAN
759	27	100.0	248	6	ABU11340	AbU11340	EST clone
760	27	100.0	248	7	ABO79110	AbO79110	Pseudomon
761	27	100.0	249	7	ABO81574	AbO81574	Pseudomon
762	27	100.0	250	4	ABG69853	AbG69853	Drosophil
763	27	100.0	250	8	ADY05200	AdY05200	Plant ful
764	27	100.0	251	8	ADX73980	AdX73980	Plant ful
765	27	100.0	252	8	ADY06653	AdY06653	Plant ful
766	27	100.0	253	8	ADX90481	AdX90481	Plant ful
767	27	100.0	254	4	AAU01403	AaU01403	Human TAN
768	27	100.0	254	4	AAU01354	AaU01354	Human TAN
769	27	100.0	254	4	AAU01404	AaU01404	Human TAN
770	27	100.0	254	4	AAU01405	AaU01405	Human TAN
771	27	100.0	254	4	AAU01406	AaU01406	Human TAN
772	27	100.0	254	8	ADT58400	AdT58400	Plant pol
773	27	100.0	254	8	ADX67034	AdX67034	Plant ful
774	27	100.0	254	8	ADX72102	AdX72102	Plant ful
775	27	100.0	255	8	ADS12267	AdS12267	Human the
776	27	100.0	256	3	AAU17737	AaU17737	Arabidops
777	27	100.0	256	8	ADG22521	AdG22521	Cyanophag
778	27	100.0	256	8	ADY05167	AdY05167	Plant ful
779	27	100.0	257	3	AAU18066	AaU18066	Arabidops
780	27	100.0	257	4	ABG59585	AbG59585	Human pro
781	27	100.0	257	6	ABC25489	AbC25489	Aspergill
782	27	100.0	257	6	ADG31429	AdG31429	Human nov
783	27	100.0	259	6	ABJ26089	AbJ26089	Aspergill
784	27	100.0	260	3	AAU14209	AaU14209	Arabidops
785	27	100.0	260	4	AAU01379	AaU01379	Human TAN
786	27	100.0	260	7	ABO76277	AbO76277	Pseudomon
787	27	100.0	260	7	ABO75253	AbO75253	Pseudomon
788	27	100.0	261	3	AAU42113	AaU42113	Human ORF
789	27	100.0	261	3	AAU44344	AaU44344	Arabidops
790	27	100.0	261	4	AAU01386	AaU01386	Human TAN
791	27	100.0	261	7	ABO70690	AbO70690	Pseudomon
792	27	100.0	261	8	ADS11930	AdS11930	Human the
793	27	100.0	261	8	ADY09025	AdY09025	Plant ful
794	27	100.0	261	8	ADY23821	AdY23821	Plant ful
795	27	100.0	262	8	ADN05438	AdN05438	Antipsori
796	27	100.0	262	8	ADX88430	AdX88430	Plant ful
797	27	100.0	263	7	ABO68160	AbO68160	Pseudomon
798	27	100.0	264	5	ABP43591	AbP43591	Human clo
799	27	100.0	265	2	AAU93156	AaU93156	Mature in
800	27	100.0	265	2	AAW46518	Aaw46518	Amino aci
801	27	100.0	265	2	AAW79128	Aaw79128	FLGA inse
802	27	100.0	265	2	AAV31242	Aav31242	Human ino
803	27	100.0	265	3	AB28524	Ab28524	Human ino
804	27	100.0	265	4	ABG04963	AbG04963	Novel hum
805	27	100.0	265	5	AAE25599	Aae25599	Human ino
806	27	100.0	265	7	ABO73861	AbO73861	Pseudomon
807	27	100.0	265	7	ABO77006	AbO77006	Pseudomon
808	27	100.0	265	8	ADJ99981	AdJ99981	Human ino
809	27	100.0	265	8	ADQ95783	AdQ95783	Human ino
810	27	100.0	265	8	ADY22546	AdY22546	Plant ful
811	27	100.0	267	8	ADX94314	AdX94314	Plant ful
812	27	100.0	268	7	ABO78949	AbO78949	Pseudomon
813	27	100.0	268	8	ADX71012	AdX71012	Plant ful
814	27	100.0	268	8	ADY09718	AdY09718	Plant ful
815	27	100.0	269	8	ADY12645	AdY12645	Plant ful
816	27	100.0	269	8	ADY06431	AdY06431	Plant ful
817	27	100.0	270	4	AAU01381	AaU01381	Human TAN
818	27	100.0	271	8	ADX96089	AdX96089	Plant ful
819	27	100.0	272	7	ABO71098	AbO71098	Pseudomon
820	27	100.0	272	8	ADP30014	AdP30014	Human sec
821	27	100.0	273	4	AB93616	Ab93616	Human pro
822	27	100.0	273	8	ADY12652	AdY12652	Plant ful
823	27	100.0	275	4	AAU42298	AaU42298	Propionib
824	27	100.0	275	6	ABM38817	AbM38817	Propionib
825	27	100.0	276	8	ADY05319	AdY05319	Plant ful
826	27	100.0	276	8	ADY04618	AdY04618	Plant ful
827	27	100.0	278	7	ABO77293	AbO77293	Pseudomon

828	27	100.0	278	8	ADQ65360	Novel hum	Adq65360	901	27	100.0	307	7	ABO73762	Pseudomon
829	27	100.0	278	8	ADY04855	Plant ful	Ady04855	902	27	100.0	308	8	ADY10226	Plant ful
830	27	100.0	280	2	AAV59722	Secreted	Aay59722	903	27	100.0	310	6	AAE36006	Human BCM
831	27	100.0	280	7	ADP77897	Human pro	Adp77897	904	27	100.0	310	7	ADC31119	Human nov
832	27	100.0	280	8	ADP19553	Human sec	Adp19553	905	27	100.0	310	7	ABO71240	Pseudomon
833	27	100.0	280	8	ADY11845	Plant ful	Ady11845	906	27	100.0	310	7	ABO68151	Pseudomon
834	27	100.0	280	9	ADZ689357	Secreted	Adz689357	907	27	100.0	310	8	ADO60536	Human NSE
835	27	100.0	281	8	ADY25023	Plant ful	Ady25023	908	27	100.0	311	3	AAI75111	Arabidops
836	27	100.0	282	8	ABM84209	Human dia	Abm84209	909	27	100.0	311	7	ABO78990	Pseudomon
837	27	100.0	282	8	ADT58503	Plant pol	Adt58503	910	27	100.0	312	7	ABO70774	Pseudomon
838	27	100.0	282	8	ADY22646	Plant ful	Ady22646	911	27	100.0	312	8	ADX73102	Plant ful
839	27	100.0	282	8	ADY11261	Plant ful	Ady11261	912	27	100.0	313	4	AAU43559	Propionib
840	27	100.0	282	8	ADX67178	Plant ful	Adx67178	913	27	100.0	313	6	ABM40078	Propionib
841	27	100.0	282	8	ADY09052	Plant ful	Ady09052	914	27	100.0	313	8	ADX76061	Plant ful
842	27	100.0	282	8	ADX67578	Plant ful	Adx67578	915	27	100.0	314	4	ABR71315	Drosophil
843	27	100.0	283	4	ABG11362	Novel hum	Abg11362	916	27	100.0	314	7	ABO70090	Pseudomon
844	27	100.0	283	7	ABO84094	Pseudomon	Abm84094	917	27	100.0	315	8	ADX91851	Plant ful
845	27	100.0	283	7	ABO70272	Pseudomon	Abm70272	918	27	100.0	315	8	ADX96015	Plant ful
846	27	100.0	284	2	AAV04777	Mycobacte	Aay04777	919	27	100.0	316	5	ABR41800	Human ova
847	27	100.0	284	3	AAI7736	Arabidops	Aag17736	920	27	100.0	316	7	ABO72619	Pseudomon
848	27	100.0	284	8	ADU16451	M. tuberc	Adul6451	921	27	100.0	316	8	ADY07871	Plant ful
849	27	100.0	285	8	ABM84208	Human dia	Abm84208	922	27	100.0	317	8	ADY05512	Plant ful
850	27	100.0	287	3	AAI7512	Arabidops	Aag17512	923	27	100.0	317	9	ADZ13501	Murine ca
851	27	100.0	287	4	AAU30555	Novel hum	Aau30555	924	27	100.0	318	6	ABU70492	Human adi
852	27	100.0	287	9	AEA20252	Novel hum	Aea20252	925	27	100.0	318	7	ABO74103	Pseudomon
853	27	100.0	288	2	AAW60676	IMP-18p m	Aaw60676	926	27	100.0	318	8	ADY12625	Plant ful
854	27	100.0	288	7	ADN95287	Human BEC	Adn95287	927	27	100.0	319	4	ABB60069	Drosophil
855	27	100.0	288	8	ADJ75575	Marker ge	Adj75575	928	27	100.0	320	6	ABM69200	Phototrab
856	27	100.0	288	8	ABM81954	Tumour-as	Abm81954	929	27	100.0	321	8	ADT54999	E. coli b
857	27	100.0	288	8	ADU06694	Novel bro	Adu06694	930	27	100.0	321	8	ADT55001	E. coli b
858	27	100.0	288	8	ADU688972	Human myo	Adu688972	931	27	100.0	321	8	ADT55002	E. coli b
859	27	100.0	289	7	ADBE60236	Rat Prote	Ade60236	932	27	100.0	321	8	ADT54989	E. coli b
860	27	100.0	289	7	ABO68736	Pseudomon	Abm68736	933	27	100.0	321	8	ADT55004	E. coli b
861	27	100.0	290	7	ABO76940	Pseudomon	Abm76940	934	27	100.0	321	8	ADT54988	E. coli b
862	27	100.0	291	4	AAU04901	Micromono	Aau04901	935	27	100.0	321	8	ADT54990	E. coli b
863	27	100.0	292	7	ABO71427	Pseudomon	Abm71427	936	27	100.0	321	8	ADT54996	E. coli b
864	27	100.0	292	8	ADY13942	Plant ful	Ady13942	937	27	100.0	321	8	ADT54980	Wild-type
865	27	100.0	292	8	ADX90335	Plant ful	Adx90335	938	27	100.0	321	8	ADT54992	E. coli b
866	27	100.0	292	8	ADX97146	Plant ful	Adx97146	939	27	100.0	321	8	ADT54995	E. coli b
867	27	100.0	293	8	ADU02343	Novel hum	Adu02343	940	27	100.0	321	8	ADT54997	E. coli b
868	27	100.0	294	2	AAI78729	Human bon	Aar78729	941	27	100.0	321	8	ADT54993	E. coli b
869	27	100.0	294	2	AAW26589	Human bon	Aaw26589	942	27	100.0	321	8	ADT55000	E. coli b
870	27	100.0	294	4	AAE10972	Human bon	Aae10972	943	27	100.0	321	8	ADT54998	E. coli b
871	27	100.0	294	4	ABO09064	Thermus c	Abb09064	944	27	100.0	321	8	ADT54986	E. coli b
872	27	100.0	294	6	ABG73289	Human bon	Abm73289	945	27	100.0	321	8	ADT55003	E. coli b
873	27	100.0	294	7	ABO76566	Pseudomon	Abm76566	946	27	100.0	321	8	ADT54994	E. coli b
874	27	100.0	294	8	ADX88323	Plant ful	Adx88323	947	27	100.0	321	8	ADT54985	E. coli b
875	27	100.0	294	8	ADY24960	Plant ful	Ady24960	948	27	100.0	321	8	ADT54991	E. coli b
876	27	100.0	295	8	ADY04881	Plant ful	Ady04881	949	27	100.0	322	2	AAI31601	Escherich
877	27	100.0	296	7	ABO68639	Pseudomon	Abm68639	950	27	100.0	323	4	ABB09053	Thermus c
878	27	100.0	296	7	ABO69874	Pseudomon	Abm69874	951	27	100.0	323	5	ABB55129	Lactococc
879	27	100.0	296	7	ABO74784	Pseudomon	Abm74784	952	27	100.0	324	7	ABO73760	Pseudomon
880	27	100.0	297	8	ADSI1813	Human the	Ads11813	953	27	100.0	325	7	ABM87136	Rice abio
881	27	100.0	297	9	AEA21066	Novel hum	Aea21066	954	27	100.0	325	8	ADY05098	Plant ful
882	27	100.0	298	8	ADG22604	Cyanophag	Adg22604	955	27	100.0	326	7	ADX79820	Bacterial
883	27	100.0	299	8	ADX91385	Plant ful	Adx91385	956	27	100.0	326	8	ADT54980	Plant ful
884	27	100.0	300	7	ABM78507	Pseudomon	Abm78507	957	27	100.0	327	7	ABO67974	Pseudomon
885	27	100.0	300	8	ABM84207	Human dia	Abm84207	958	27	100.0	327	7	ABM88488	Rice abio
886	27	100.0	300	8	ADU01159	Human pro	Adu01159	959	27	100.0	328	7	ABO83072	Pseudomon
887	27	100.0	300	8	ADU15541	Novel hum	Adu15541	960	27	100.0	328	7	ABO65283	Klebsiell
888	27	100.0	301	7	ABO68659	Pseudomon	Abm68659	961	27	100.0	330	7	ADC33088	Human nov
889	27	100.0	301	7	ABO78495	Pseudomon	Abm78495	962	27	100.0	331	5	AAE25292	Human nuc
890	27	100.0	301	8	ADX77511	Plant ful	Adx77511	963	27	100.0	332	7	ABM88586	Rice abio
891	27	100.0	301	8	ADY05292	Plant ful	Ady05292	964	27	100.0	332	9	ABM90849	M. xanthu
892	27	100.0	302	4	AAE88546	Haemophil	Aae88546	965	27	100.0	333	4	ABG19326	Novel hum
893	27	100.0	302	5	AAU91477	Haemophil	Aau91477	966	27	100.0	333	7	ABO79386	Pseudomon
894	27	100.0	302	7	ABO75296	Pseudomon	Abm75296	967	27	100.0	335	2	AAW97110	Thermosta
895	27	100.0	303	3	AAE54288	Human pan	Aab54288	968	27	100.0	335	8	ADI42738	Plant tra
896	27	100.0	305	4	AAE95872	Human pro	Abm95872	969	27	100.0	336	8	ADX93937	Plant ful
897	27	100.0	305	7	ABO68848	Pseudomon	Abm68848	970	27	100.0	336	8	ADX96963	Plant ful
898	27	100.0	305	8	ADN96683	Novel hum	Adn96683	971	27	100.0	338	4	ABB61114	Drosophil
899	27	100.0	305	8	ADY22587	Plant ful	Ady22587	972	27	100.0	338	7	ABO79800	Pseudomon
900	27	100.0	305	9	ADW17657	Pinus rad	Adw17657	973	27	100.0	338	7	ABO74796	Pseudomon

974 27 100.0 338 8 ADY04253 Plant ful  
975 27 100.0 339 8 ADY07123 Plant ful  
976 27 100.0 340 4 AAB92631 Human pro  
977 27 100.0 340 8 ADI82535 Human mod  
978 27 100.0 340 8 ADY12982 Plant ful  
979 27 100.0 342 7 ABO83285 Pseudomon  
980 27 100.0 344 7 ABG75045 Method of  
981 27 100.0 344 7 ABO73740 Pseudomon  
982 27 100.0 346 7 ABO78421 Pseudomon  
983 27 100.0 346 8 ADX76966 Plant ful  
984 27 100.0 347 8 ADY05455 Plant ful  
985 27 100.0 347 9 ABM94214 M. xanthu  
986 27 100.0 348 7 ABO77897 Pseudomon  
987 27 100.0 350 7 ABO76813 Pseudomon  
988 27 100.0 352 3 AAG14208 Arabidops  
989 27 100.0 352 8 ADQ66999 Novel hum  
990 27 100.0 353 8 ADG97881 Mouse can  
991 27 100.0 354 7 ADC31486 Human nov  
992 27 100.0 355 8 ADJ70245 Mouse Nur  
993 27 100.0 356 8 ADX68112 Plant ful  
994 27 100.0 358 8 ADX91609 Plant ful  
995 27 100.0 358 8 ADY10139 Plant ful  
996 27 100.0 359 3 AAG29857 Arabidops  
997 27 100.0 359 7 ABM87580 Rice abio  
998 27 100.0 359 8 ABO84931 Murine ca  
999 27 100.0 359 8 ADY10486 Plant ful  
1000 27 100.0 360 7 ABO68805 Pseudomon

ALIGNMENTS

RESULT 1  
ID AAW86180 standard; peptide; 6 AA.  
XX  
AC AAW86180;  
XX  
DT 04-MAR-1999 (first entry)  
XX  
DE Peptide used in a method for inhibiting wound contraction.  
XX  
KW wound contraction; reduction; inhibition; tissue regeneration; scar;  
KW wound; joint motion; body deformation.  
XX  
OS Synthetic.  
XX  
PN US5851994-A.  
XX  
PD 22-DEC-1998.  
XX  
PF 06-JUN-1995; 95US-00473025.  
XX  
PR 28-APR-1994; 94US-00234979.  
XX  
PA (LJOL-) LA JOLLA CANCER RES FOUND.  
XX  
PI Polarek J, Schreiber R;  
XX  
DR WPI; 1999-080478/07.  
XX  
PT Inhibition of wound contraction - with peptide derivatives rich in basic  
PT amino acids.  
XX  
PS Claim 11; Col 24; 16pp; English.  
XX  
CC The invention provides methods for reduction or inhibition of wound  
CC contraction that comprises administration of a peptide having more than 3  
CC consecutive basic amino acid residues. Alternatively, the peptide  
CC contains the amino acid sequence Arg-Gly-Asp and a basic amino acid  
CC sequence, or the peptide comprises 6-30 amino acids in which at least 4  
CC out of a sequence of 6 consecutive amino acids are basic amino acids. The  
CC method is used to allow normal tissue regeneration without excessive scar

CC formation which, in the case of large wounds, can result in loss of joint  
CC motion or major body deformation. Sequences AAW86170 to AAW86183  
CC represent specifically claimed examples of peptides that can be used in  
CC the method of the invention  
XX  
SQ Sequence 6 AA;  
Query Match 100.0%; Score 27; DB 2; Length 6;  
Best Local Similarity 100.0%; Pred. No. 2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GRRGR 5  
Db |||||  
1 GRRGR 5  
RESULT 2  
ABP15087  
ID ABP15087 standard; peptide; 7 AA.  
XX  
AC ABP15087;  
XX  
DT 11-SEP-2003 (revised)  
DT 15-JUL-2002 (first entry)  
XX  
DE HIV A03 super motif vpr peptide #11.  
XX  
KW HIV; HIV-1; human immunodeficiency virus; env; pol; gag; nef; vpr; vpu;  
KW vif; tat; cytotoxic T lymphocyte; CTL; immune response; epitope; antigen;  
KW vaccine; HIV infection; immunisation; virucide.  
XX  
OS Human immunodeficiency virus 1.  
XX  
PN WO200124810-A1.  
XX  
PD 12-APR-2001.  
XX  
PF 05-OCT-2000; 2000WO-US027766.  
XX  
PR 05-OCT-1999; 99US-00412863.  
XX  
PA (EPIM-) EPIMUNE INC.  
XX  
PI Sette A, Sidney J, Southwood S, Livingston BD, Chesnut R;  
PI Baker DM, Celis E, Kubo RT, Grey HM;  
XX  
DR WPI; 2001-354887/37.  
XX  
PT Vaccine compositions comprising human immunodeficiency virus-1 (HIV-1)  
PT peptide groups, useful for vaccinating against HIV-1.  
XX  
PS Claim 32; Page 177; 448pp; English.  
XX  
CC The present invention describes a composition (I) comprising a prepared  
CC human immunodeficiency virus-1 (HIV-1) group comprising an amino acid  
CC sequence selected from 51 defined amino acid sequences (ABL25347 to  
CC ABP25397). (I) has virucide activity and can be used in vaccines. (I) may  
CC be used for immunising subjects against HIV-1 infections. The use of  
CC group-based vaccines has several advantages over traditional vaccines,  
CC particularly when compared to the use of whole antigens in vaccine  
CC compositions. There is evidence that the immune response to whole  
CC antigens is directed largely toward variable regions of the antigen,  
CC allowing for immune escape due to mutations. The groups for inclusion in  
CC an group-based vaccine may be selected from conserved regions of viral or  
CC tumour-associated antigens, which therefore reduces the likelihood of  
CC escape mutants. Furthermore, immunosuppressive groups that may be present  
CC in whole antigens can be avoided with the use of group-based vaccines. An  
CC additional advantage of an group-based vaccine approach is the ability to  
CC combine selected groups (CTL and HTL), and further, to modify the  
CC composition of the groups, achieving, for example, enhanced  
CC immunogenicity. Accordingly, the immune response can be modulated, as  
CC appropriate, for the target disease. Similar engineering of the response  
CC is not possible with traditional approaches. ABP11501 to ABP25412

CC represent peptide sequences used in the exemplification of the present  
 CC invention. (Updated on 11-SEP-2003 to standardise OS field)

XX  
 SQ Sequence 7 AA;

Query Match 100.0%; Score 27; DB 4; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 2e+06; 0; Gaps 0;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
 Db |||||  
 3 GRRGR 7

RESULT 3  
 ABP22159  
 ID ABP22159 standard; peptide; 7 AA.  
 XX  
 AC ABP22159;  
 XX  
 DT 11-SEP-2003 (revised)  
 DT 15-JUL-2002 (first entry)  
 XX  
 DE HIV A03 motif vpr peptide #1.  
 XX  
 KW HIV; HIV-1; human immunodeficiency virus; env; pol; gag; nef; vpr; vpu;  
 KW vif; tat; cytotoxic T lymphocyte; CTL; immune response; epitope; antigen;  
 KW vaccine; HIV infection; immunisation; virucide.  
 XX  
 OS Human immunodeficiency virus 1.  
 XX  
 PN WO200124810-A1.  
 XX  
 PD 12-APR-2001.  
 XX  
 PP 05-OCT-2000; 2000WO-US027766.  
 XX  
 PR 05-OCT-1999; 99US-00412863.  
 XX  
 PA (EPIM-) EPIMMUNE INC.  
 XX  
 PI Sette A, Sidney J, Southwood S, Livingston BD, Cheanut R;  
 PI Baker DM, Cellis E, Kubo RT, Grey HM;  
 XX  
 DR WPI; 2001-354887/37.  
 XX  
 PT Vaccine compositions comprising human immunodeficiency virus-1 (HIV-1)  
 PT peptide groups, useful for vaccinating against HIV-1.  
 XX  
 PS Claim 32; Page 323; 448pp; English.

CC The present invention describes a composition (I) comprising a prepared  
 CC human immunodeficiency virus-1 (HIV-1) group comprising an amino acid  
 CC sequence selected from 51 defined amino acid sequences (ABL25347 to  
 CC ABP25397). (I) has virucide activity and can be used in vaccines. (I) may  
 CC be used for immunising subjects against HIV-1 infections. The use of  
 CC group-based vaccines has several advantages over traditional vaccines,  
 CC particularly when compared to the use of whole antigens in vaccine  
 CC compositions. There is evidence that the immune response to whole  
 CC antigens is directed largely toward variable regions of the antigen,  
 CC allowing for immune escape due to mutations. The groups for inclusion in  
 CC an group-based vaccine may be selected from conserved regions of viral or  
 CC tumour-associated antigens, which therefore reduces the likelihood of  
 CC escape mutants. Furthermore, immunosuppressive groups that may be present  
 CC in whole antigens can be avoided with the use of group-based vaccines. An  
 CC additional advantage of an group-based vaccine approach is the ability to  
 CC combine selected groups (CTL and HTL), and further, to modify the  
 CC composition of the groups, achieving, for example, enhanced  
 CC immunogenicity. Accordingly, the immune response can be modulated, as  
 CC appropriate, for the target disease. Similar engineering of the response  
 CC is not possible with traditional approaches. ABP11501 to ABP25412  
 CC represent peptide sequences used in the exemplification of the present  
 CC invention. (Updated on 11-SEP-2003 to standardise OS field)

XX  
 SQ Sequence 7 AA;

Query Match 100.0%; Score 27; DB 4; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 2e+06; 0; Gaps 0;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
 Db |||||  
 3 GRRGR 7

RESULT 4  
 ADG44442  
 ID ADG44442 standard; peptide; 7 AA.  
 XX  
 AC ADG44442;  
 XX  
 DT 26-FEB-2004 (first entry)  
 XX  
 DE Anti-pathogen related amino acid sequence SEQ ID NO:241.  
 XX  
 KW pathogen infection; chimeric molecule; pathogen-detection domain;  
 KW pathogen-induced product-detection domain; effector domain;  
 KW antimicrobial; gene therapy.  
 XX  
 OS Synthetic.  
 XX  
 PN WO2003066003-A2.  
 XX  
 PD 14-AUG-2003.  
 XX  
 PF 07-FEB-2003; 2003WO-US003978.  
 XX  
 PR 07-FEB-2002; 2002US-0355022P.  
 PR 07-FEB-2002; 2002US-0355359P.  
 PR 10-DEC-2002; 2002US-0432386P.  
 XX  
 PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
 XX  
 PI Rider TH;  
 XX  
 DR WPI; 2003-731477/59.  
 XX  
 PT Treating or preventing a pathogen infection in a cell by administering to  
 PT the cell chimeric molecules having a pathogen-detection domain or  
 PT pathogen-induced product-detection domain and at least one effector  
 PT domain.  
 XX  
 PS Example 7; SEQ ID NO 241; 348pp; English.

CC The present invention describes a method for treating or preventing a  
 CC pathogen infection in a cell. The method comprises administering to the  
 CC cell chimeric molecules (I) having at least one pathogen-detection domain  
 CC or pathogen-induced product-detection domain and at least one effector  
 CC domain. The pathogen-detection domain or pathogen-induced product-  
 CC detection domain is not naturally bound to the effector domain. In the  
 CC presence of a pathogen or pathogen-induced product in the cell, the  
 CC chimeric molecules bind to the pathogen and activate the effector domain.  
 CC Also described: (1) treating or preventing the spread of a pathogen  
 CC infection in an organism; (2) a chimeric molecule (I) having at least one  
 CC pathogen-detection domain or pathogen-induced product-detection domain  
 CC and at least one effector domain; (3) an agent having at least one  
 CC pathogen-interacting or pathogen-induced product-interacting molecular  
 CC structure and at least one effector-mediated molecular structure; and  
 CC (4) an assay for detecting a pathogen infection in a cell or organism.  
 CC (I) has antimicrobial activity, and can be used in gene therapy. The  
 CC method is useful for treating or preventing a pathogen infection in a  
 CC cell or organism. The present sequence is used in the exemplification of  
 CC the present invention.

XX  
 SQ Sequence 7 AA;

Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
RESULT 5					
ADG44571					
ID	ADG44571	standard;	peptide;	7 AA.	
XX					
AC	ADG44571;				
DT	26-FEB-2004	(first entry)			
XX					
DE	Anti-pathogen related amino acid sequence SEQ ID NO:370.				
XX					
KW	pathogen infection; chimeric molecule; pathogen-detection domain;				
KW	pathogen-induced product-detection domain; effector domain;				
KW	antimicrobial; gene therapy.				
XX					
OS	Synthetic.				
XX					
PN	WO2003066003-A2.				
XX					
PD	14-AUG-2003.				
XX					
PF	07-FEB-2003;	2003WO-US003978.			
XX					
PR	07-FEB-2002;	2002US-0355022P.			
XX					
PR	07-FEB-2002;	2002US-0355359P.			
XX					
PR	10-DEC-2002;	2002US-0432386P.			
XX					
PA	(MASI ) MASSACHUSETTS INST TECHNOLOGY.				
XX					
PI	Rider TH;				
XX					
DR	WPI; 2003-731477/69.				
XX					
PT	Treating or preventing a pathogen infection in a cell by administering to				
PT	the cell chimeric molecules having a pathogen-detection domain or				
PT	pathogen-induced product-detection domain and at least one effector				
PT	domain.				
PS	Example 7; SEQ ID NO 370;	348pp; English.			
XX					
CC	The present invention describes a method for treating or preventing a				
CC	pathogen infection in a cell. The method comprises administering to the				
CC	cell chimeric molecules (1) having at least one pathogen-detection domain				
CC	or pathogen-induced product-detection domain and at least one effector				
CC	domain. The pathogen-detection domain or pathogen-induced product-				
CC	detection domain is not naturally bound to the effector domain. In the				
CC	presence of a pathogen or pathogen-induced product in the cell, the				
CC	chimeric molecules bind to the pathogen and activate the effector domain.				
CC	Also described: (1) treating or preventing the spread of a pathogen				
CC	infection in an organism; (2) a chimeric molecule (1) having at least one				
CC	pathogen-detection domain or pathogen-induced product-detection domain				
CC	and at least one effector domain; (3) an agent having at least one				
CC	pathogen-interacting or pathogen-induced product-interacting molecular				
CC	structure and at least one effector-mediated molecular structure; and				
CC	(4) an assay for detecting a pathogen infection in a cell or organism.				
CC	(1) has antimicrobial activity, and can be used in gene therapy. The				
CC	method is useful for treating or preventing a pathogen infection in a				
CC	cell or organism. The present sequence is used in the exemplification of				
CC	the present invention.				
XX	Sequence 7 AA;				
SQ					
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	Length 7;
Best Local Similarity		100.0%;	Pred. No. 2e+06;	0;	Indels 0;
Matches		5;	Conservative	0;	Mismatches 0;
Qy	1 GRRGR 5				
Db	1 GRRGR 5				
Query Match		100.0%;	Score 27;	DB 7;	

```

Db          1 GRGR 5

RESULT 7
ADG44543
ID  ADG44543 standard; peptide; 7 AA.
XX
AC  ADG44543;
XX
DT  26-FEB-2004 (first entry)
XX
DE  Anti-pathogen related amino acid sequence SEQ ID NO:342.
XX
KW  pathogen infection; chimeric molecule; pathogen-detection domain;
KW  pathogen-induced product-detection domain; effector domain;
KW  antimicrobial; gene therapy.
XX
OS  Synthetic.
XX
PN  WO2003066003-A2.
XX
PD  14-AUG-2003.
XX
PF  07-FEB-2003; 2003WO-US003978.
XX
PR  07-FEB-2002; 2002US-0355022P.
PR  07-FEB-2002; 2002US-0355359P.
PR  10-DEC-2002; 2002US-0432386P.
XX
PA  (MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX
PI  Rider TH;
XX
WPI; 2003-731477/69.
XX
Treating or preventing a pathogen infection in a cell by administering to
the cell chimeric molecules having a pathogen-detection domain or
pathogen-induced product-detection domain and at least one effector
domain.
XX
Example 7; SEQ ID NO 342; 348pp; English.
XX
The present invention describes a method for treating or preventing a
pathogen infection in a cell. The method comprises administering to the
cell chimeric molecules (I) having at least one pathogen-detection domain
or pathogen-induced product-detection domain and at least one effector
domain. The pathogen-detection domain or pathogen-induced product-
detection domain is not naturally bound to the effector domain. In the
presence of a pathogen or pathogen-induced product in the cell, the
chimeric molecules bind to the pathogen and activate the effector domain.
Also described: (1) treating or preventing the spread of a pathogen
infection in an organism; (2) a chimeric molecule (I) having at least one
pathogen-detection domain or pathogen-induced product-detection domain
and at least one effector domain; (3) an agent having at least one
pathogen-interacting or pathogen-induced product-interacting molecular
structure and at least one effector-mediated molecular structure; and
(4) an assay for detecting a pathogen infection in a cell or organism.
The (I) has antimicrobial activity, and can be used in gene therapy. The
method is useful for treating or preventing a pathogen infection in a
cell or organism. The present sequence is used in the exemplification of
the present invention.
XX
Sequence 7 AA;
XX
Query Match          100.0%; Score 27; DB 7; Length 7;
Best Local Similarity 100.0%; Pred. No. 2e+06;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY          1 GRGR 5
          |||||
Db          1 GRGR 5

RESULT 9
ADG44515
ID  ADG44515 standard; peptide; 7 AA.
XX
AC  ADG44515;
XX
DT  26-FEB-2004 (first entry)
XX
DE  Anti-pathogen related amino acid sequence SEQ ID NO:454.
XX
KW  pathogen infection; chimeric molecule; pathogen-detection domain;
KW  pathogen-induced product-detection domain; effector domain;
KW  antimicrobial; gene therapy.
XX
OS  Synthetic.
XX
PN  WO2003066003-A2.
XX
PD  14-AUG-2003.
XX
PF  07-FEB-2003; 2003WO-US003978.
XX
PR  07-FEB-2002; 2002US-0355022P.
PR  07-FEB-2002; 2002US-0355359P.
PR  10-DEC-2002; 2002US-0432386P.
XX
PA  (MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX
PI  Rider TH;
XX
WPI; 2003-731477/69.
XX
Treating or preventing a pathogen infection in a cell by administering to
the cell chimeric molecules having a pathogen-detection domain or
pathogen-induced product-detection domain and at least one effector
domain.
XX
Example 7; SEQ ID NO 454; 348pp; English.
XX
The present invention describes a method for treating or preventing a
pathogen infection in a cell. The method comprises administering to the
cell chimeric molecules (I) having at least one pathogen-detection domain
or pathogen-induced product-detection domain and at least one effector
domain. The pathogen-detection domain or pathogen-induced product-
detection domain is not naturally bound to the effector domain. In the
presence of a pathogen or pathogen-induced product in the cell, the
chimeric molecules bind to the pathogen and activate the effector domain.
Also described: (1) treating or preventing the spread of a pathogen
infection in an organism; (2) a chimeric molecule (I) having at least one
pathogen-detection domain or pathogen-induced product-detection domain
and at least one effector domain; (3) an agent having at least one
pathogen-interacting or pathogen-induced product-interacting molecular
structure and at least one effector-mediated molecular structure; and
(4) an assay for detecting a pathogen infection in a cell or organism.
The (I) has antimicrobial activity, and can be used in gene therapy. The
method is useful for treating or preventing a pathogen infection in a
cell or organism. The present sequence is used in the exemplification of
the present invention.
XX
Sequence 7 AA;
XX
Query Match          100.0%; Score 27; DB 7; Length 7;
Best Local Similarity 100.0%; Pred. No. 2e+06;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY          1 GRGR 5
          |||||
Db          1 GRGR 5

RESULT 8
ADG44655
ID  ADG44655 standard; peptide; 7 AA.
XX
AC  ADG44655;
XX
DT  26-FEB-2004 (first entry)
XX
DE  Anti-pathogen related amino acid sequence SEQ ID NO:454.
XX
KW  pathogen infection; chimeric molecule; pathogen-detection domain;
KW  pathogen-induced product-detection domain; effector domain;
KW  antimicrobial; gene therapy.
XX
OS  Synthetic.
XX
PN  WO2003066003-A2.
XX
PD  14-AUG-2003.
XX
PF  07-FEB-2003; 2003WO-US003978.
XX
PR  07-FEB-2002; 2002US-0355022P.
PR  07-FEB-2002; 2002US-0355359P.
PR  10-DEC-2002; 2002US-0432386P.
XX
PA  (MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX
PI  Rider TH;
XX
WPI; 2003-731477/69.
XX
Treating or preventing a pathogen infection in a cell by administering to
the cell chimeric molecules having a pathogen-detection domain or
pathogen-induced product-detection domain and at least one effector
domain.
XX
Example 7; SEQ ID NO 454; 348pp; English.
XX
The present invention describes a method for treating or preventing a
pathogen infection in a cell. The method comprises administering to the
cell chimeric molecules (I) having at least one pathogen-detection domain
or pathogen-induced product-detection domain and at least one effector
domain. The pathogen-detection domain or pathogen-induced product-
detection domain is not naturally bound to the effector domain. In the
presence of a pathogen or pathogen-induced product in the cell, the
chimeric molecules bind to the pathogen and activate the effector domain.
Also described: (1) treating or preventing the spread of a pathogen
infection in an organism; (2) a chimeric molecule (I) having at least one
pathogen-detection domain or pathogen-induced product-detection domain
and at least one effector domain; (3) an agent having at least one
pathogen-interacting or pathogen-induced product-interacting molecular
structure and at least one effector-mediated molecular structure; and
(4) an assay for detecting a pathogen infection in a cell or organism.
The (I) has antimicrobial activity, and can be used in gene therapy. The
method is useful for treating or preventing a pathogen infection in a
cell or organism. The present sequence is used in the exemplification of
the present invention.
XX
Sequence 7 AA;
XX
Query Match          100.0%; Score 27; DB 7; Length 7;
Best Local Similarity 100.0%; Pred. No. 2e+06;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY          1 GRGR 5
          |||||
Db          1 GRGR 5

RESULT 9
ADG44515
ID  ADG44515 standard; peptide; 7 AA.

```

XX AC ADG44515;  
XX DT  
XX DT 26-FEB-2004 (first entry)  
XX DE  
XX DE Anti-pathogen related amino acid sequence SEQ ID NO:314.  
XX KW pathogen infection; chimeric molecule; pathogen-detection domain;  
XX KW pathogen-induced product-detection domain; effector domain;  
XX KW antimicrobial; gene therapy.  
XX OS Synthetic.  
XX PN WO2003066003-A2.  
XX PD 14-AUG-2003.  
XX PF 07-FEB-2003; 2003WO-US003978.  
XX PR 07-FEB-2002; 2002US-0355022P.  
XX PR 07-FEB-2002; 2002US-0355359P.  
XX PR 10-DEC-2002; 2002US-0432386P.  
XX PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
XX PI Rider TH;  
XX PI WPI; 2003-731477/69.  
XX DR Treating or preventing a pathogen infection in a cell by administering to  
XX PT the cell chimeric molecules having a pathogen-detection domain or  
XX PT pathogen-induced product-detection domain and at least one effector  
XX PT domain.  
XX PS Example 7; SEQ ID NO 314; 348pp; English.  
XX CC The present invention describes a method for treating or preventing a  
XX CC pathogen infection in a cell. The method comprises administering to the  
XX CC cell chimeric molecules (I) having at least one pathogen-detection domain  
XX CC or pathogen-induced product-detection domain and at least one effector  
XX CC domain. The pathogen-detection domain or pathogen-induced product-  
XX CC detection domain is not naturally bound to the effector domain. In the  
XX CC presence of a pathogen or pathogen-induced product in the cell, the  
XX CC chimeric molecules bind to the pathogen and activate the effector domain.  
XX CC Also described: (1) treating or preventing the spread of a pathogen  
XX CC infection in an organism; (2) a chimeric molecule (I) having at least one  
XX CC pathogen-detection domain or pathogen-induced product-detection domain  
XX CC and at least one effector domain; (3) an agent having at least one  
XX CC pathogen-interacting or pathogen-induced product-interacting molecular  
XX CC structure and at least one effector-mediated molecular structure; and  
XX CC (4) an assay for detecting a pathogen infection in a cell or organism.  
XX CC (I) has antimicrobial activity, and can be used in gene therapy. The  
XX CC method is useful for treating or preventing a pathogen infection in a  
XX CC cell or organism. The present sequence is used in the exemplification of  
XX CC the present invention.  
XX SQ Sequence 7 AA;  
Query Match 100.0%; Score 27; DB 7; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GRRGR 5  
Db 1 GRRGR 5  
RESULT 10  
ADG44669  
ID ADG44669 standard; peptide; 7 AA.  
XX AC ADG44669;  
XX AC

DT 26-FEB-2004 (first entry)  
XX Anti-pathogen related amino acid sequence SEQ ID NO:468.  
DE pathogen infection; chimeric molecule; pathogen-detection domain;  
XX pathogen-induced product-detection domain; effector domain;  
KW antimicrobial; gene therapy.  
XX Synthetic.  
XX OS WO2003066003-A2.  
XX PN 14-AUG-2003.  
XX PD 07-FEB-2003; 2003WO-US003978.  
XX PF 07-FEB-2002; 2002US-0355022P.  
XX PR 07-FEB-2002; 2002US-0355359P.  
XX PR 10-DEC-2002; 2002US-0432386P.  
XX PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
XX PI Rider TH;  
XX PI WPI; 2003-731477/69.  
XX DR Treating or preventing a pathogen infection in a cell by administering to  
XX PT the cell chimeric molecules having a pathogen-detection domain or  
XX PT pathogen-induced product-detection domain and at least one effector  
XX PT domain.  
XX PS Example 7; SEQ ID NO 468; 348pp; English.  
XX CC The present invention describes a method for treating or preventing a  
XX CC pathogen infection in a cell. The method comprises administering to the  
XX CC cell chimeric molecules (I) having at least one pathogen-detection domain  
XX CC or pathogen-induced product-detection domain and at least one effector  
XX CC domain. The pathogen-detection domain or pathogen-induced product-  
XX CC detection domain is not naturally bound to the effector domain. In the  
XX CC presence of a pathogen or pathogen-induced product in the cell, the  
XX CC chimeric molecules bind to the pathogen and activate the effector domain.  
XX CC Also described: (1) treating or preventing the spread of a pathogen  
XX CC infection in an organism; (2) a chimeric molecule (I) having at least one  
XX CC pathogen-detection domain or pathogen-induced product-interacting molecular  
XX CC structure and at least one effector-mediated molecular structure; and  
XX CC (4) an assay for detecting a pathogen infection in a cell or organism.  
XX CC (I) has antimicrobial activity, and can be used in gene therapy. The  
XX CC method is useful for treating or preventing a pathogen infection in a  
XX CC cell or organism. The present sequence is used in the exemplification of  
XX CC the present invention.  
XX SQ Sequence 7 AA;  
Query Match 100.0%; Score 27; DB 7; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GRRGR 5  
Db 1 GRRGR 5  
RESULT 11  
ADG44459  
ID ADG44459 standard; peptide; 7 AA.  
XX AC ADG44459;  
XX AC 26-FEB-2004 (first entry)  
XX DT  
XX DE Anti-pathogen related amino acid sequence SEQ ID NO:258.  
DE



XX pathogen infection; chimeric molecule; pathogen-detection domain;  
 KW pathogen-induced product-detection domain; effector domain;  
 KW antimicrobial; gene therapy.  
 XX Synthetic.  
 OS  
 XX WO2003066003-A2.  
 PN  
 XX 14-AUG-2003.  
 PD  
 XX 07-FEB-2003; 2003WO-US003978.  
 PF  
 XX 07-FEB-2002; 2002US-0355022P.  
 PR  
 XX 07-FEB-2002; 2002US-0355359P.  
 PR  
 XX 10-DEC-2002; 2002US-0432386P.  
 PR  
 XX (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
 PA  
 XX Rider TH;  
 XX WPI; 2003-731477/69.  
 PI  
 XX  
 DR Treating or preventing a pathogen infection in a cell by administering to  
 XX the cell chimeric molecules having a pathogen-detection domain or  
 PT pathogen-induced product-detection domain and at least one effector  
 PT domain.  
 PT  
 XX Example 7; SEQ ID NO 258; 348pp; English.  
 PS  
 XX The present invention describes a method for treating or preventing a  
 CC pathogen infection in a cell. The method comprises administering to the  
 CC cell chimeric molecules (I) having at least one pathogen-detection domain  
 CC or pathogen-induced product-detection domain and at least one effector  
 CC domain. The pathogen-detection domain or pathogen-induced product-  
 CC detection domain is not naturally bound to the effector domain. In the  
 CC presence of a pathogen or pathogen-induced product in the cell, the  
 CC chimeric molecules bind to the pathogen and activate the spread of a pathogen  
 CC Also described: (1) treating or preventing the spread of a pathogen  
 CC infection in an organism; (2) a chimeric molecule (I) having at least one  
 CC pathogen-detection domain or pathogen-induced product-detection domain  
 CC and at least one effector domain; (3) an agent having at least one  
 CC pathogen-interacting or pathogen-induced product-interacting molecular  
 CC structure and at least one effector-mediated molecular structure; and  
 CC (4) an assay for detecting a pathogen infection in a cell or organism.  
 CC (I) has antimicrobial activity, and can be used in gene therapy. The  
 CC method is useful for treating or preventing a pathogen infection in a  
 CC cell or organism. The present sequence is used in the exemplification of  
 CC the present invention.  
 XX  
 SQ Sequence 7 AA;  
 Query Match 100.0%; Score 27; DB 7; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 2e+06; 0; Indels 0; Gaps 0;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GRRGR 5  
 |||||  
 Db 1 GRRGR 5  
 RESULT 12  
 ADG44557  
 ID ADG44557 standard; peptide; 7 AA.  
 AC  
 XX ADG44557;  
 AC  
 XX 26-FEB-2004 (first entry)  
 DT  
 XX Anti-pathogen related amino acid sequence SEQ ID NO:356.  
 DE pathogen infection; chimeric molecule; pathogen-detection domain;  
 XX pathogen-induced product-detection domain; effector domain;  
 KW pathogen infection; chimeric molecule; pathogen-detection domain;  
 KW pathogen-induced product-detection domain; effector domain;  
 XX antimicrobial; gene therapy.  
 OS Synthetic.

XX antimicrobial; gene therapy.  
 XX Synthetic.  
 OS  
 XX WO2003066003-A2.  
 PN  
 XX 14-AUG-2003.  
 PD  
 XX 07-FEB-2003; 2003WO-US003978.  
 PF  
 XX 07-FEB-2002; 2002US-0355022P.  
 PR  
 XX 07-FEB-2002; 2002US-0355359P.  
 PR  
 XX 10-DEC-2002; 2002US-0432386P.  
 PR  
 XX (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
 PA  
 XX Rider TH;  
 XX WPI; 2003-731477/69.  
 PI  
 XX  
 DR Treating or preventing a pathogen infection in a cell by administering to  
 XX the cell chimeric molecules having a pathogen-detection domain or  
 PT pathogen-induced product-detection domain and at least one effector  
 PT domain.  
 PT  
 XX Example 7; SEQ ID NO 356; 348pp; English.  
 PS  
 XX The present invention describes a method for treating or preventing a  
 CC pathogen infection in a cell. The method comprises administering to the  
 CC cell chimeric molecules (I) having at least one pathogen-detection domain  
 CC or pathogen-induced product-detection domain and at least one effector  
 CC domain. The pathogen-detection domain or pathogen-induced product-  
 CC detection domain is not naturally bound to the effector domain. In the  
 CC presence of a pathogen or pathogen-induced product in the cell, the  
 CC chimeric molecules bind to the pathogen and activate the spread of a pathogen  
 CC Also described: (1) treating or preventing the spread of a pathogen  
 CC infection in an organism; (2) a chimeric molecule (I) having at least one  
 CC pathogen-detection domain or pathogen-induced product-detection domain  
 CC and at least one effector domain; (3) an agent having at least one  
 CC pathogen-interacting or pathogen-induced product-interacting molecular  
 CC structure and at least one effector-mediated molecular structure; and  
 CC (4) an assay for detecting a pathogen infection in a cell or organism.  
 CC (I) has antimicrobial activity, and can be used in gene therapy. The  
 CC method is useful for treating or preventing a pathogen infection in a  
 CC cell or organism. The present sequence is used in the exemplification of  
 CC the present invention.  
 XX  
 SQ Sequence 7 AA;  
 Query Match 100.0%; Score 27; DB 7; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 2e+06; 0; Indels 0; Gaps 0;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GRRGR 5  
 |||||  
 Db 1 GRRGR 5  
 RESULT 13  
 ADG44641  
 ID ADG44641 standard; peptide; 7 AA.  
 AC  
 XX ADG44641;  
 AC  
 XX 26-FEB-2004 (first entry)  
 DT  
 XX Anti-pathogen related amino acid sequence SEQ ID NO:440.  
 DE pathogen infection; chimeric molecule; pathogen-detection domain;  
 XX pathogen-induced product-detection domain; effector domain;  
 KW pathogen infection; chimeric molecule; pathogen-detection domain;  
 KW pathogen-induced product-detection domain; effector domain;  
 XX antimicrobial; gene therapy.  
 OS Synthetic.

XX PN WO2003066003-A2.  
XX PD 14-AUG-2003.  
XX PF 07-FEB-2003; 2003WO-US003978.  
XX PR 07-FEB-2002; 2002US-0355022P.  
XX PR 07-FEB-2002; 2002US-0355359P.  
XX PR 10-DEC-2002; 2002US-0432386P.  
XX PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
XX PI Rider TH;  
XX DR WPI; 2003-731477/69.  
XX XX Treating or preventing a pathogen infection in a cell by administering to  
PT the cell chimeric molecules having a pathogen-detection domain or  
PT pathogen-induced product-detection domain and at least one effector  
PT domain.  
XX PS Example 7; SEQ ID NO 440; 348pp; English.  
XX CC The present invention describes a method for treating or preventing a  
CC pathogen infection in a cell. The method comprises administering to the  
CC cell chimeric molecules (I) having at least one pathogen-detection domain  
CC or pathogen-induced product-detection domain and at least one effector  
CC domain. The pathogen-detection domain or pathogen-induced product-  
CC detection domain is not naturally bound to the effector domain. In the  
CC presence of a pathogen or pathogen-induced product in the cell, the  
CC chimeric molecules bind to the pathogen and activate the spread of a pathogen  
CC infection in an organism; (2) a chimeric molecule (I) having at least one  
CC pathogen-detection domain or pathogen-induced product-detection domain  
CC and at least one effector domain; (3) an agent having at least one  
CC pathogen-interacting or pathogen-induced product-interacting molecular  
CC structure and at least one effector-mediated product-interacting molecular  
CC (4) an assay for detecting a pathogen infection in a cell or organism.  
CC (I) has antimicrobial activity, and can be used in gene therapy. The  
CC method is useful for treating or preventing a pathogen infection in a  
CC cell or organism. The present sequence is used in the exemplification of  
XX the present invention.  
XX SQ Sequence 7 AA;  
Query Match 100.0%; Score 27; DB 7; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GRRGR 5  
Db |||||  
1 GRRGR 5  
RESULT 14  
ADG44473  
ID ADG44473 standard; peptide; 7 AA.  
XX AC ADG44473;  
XX DT 26-FEB-2004 (first entry)  
XX DE Anti-pathogen related amino acid sequence SEQ ID NO:272.  
XX KW pathogen infection; chimeric molecule; pathogen-detection domain;  
KW pathogen-induced product-detection domain; effector domain;  
KW antimicrobial; gene therapy.  
XX OS Synthetic.  
XX PN WO2003066003-A2.  
XX PF 07-FEB-2003; 2003WO-US003978.

PD XX 14-AUG-2003.  
XX PF 07-FEB-2003; 2003WO-US003978.  
XX PR 07-FEB-2002; 2002US-0355022P.  
XX PR 07-FEB-2002; 2002US-0355359P.  
XX PR 10-DEC-2002; 2002US-0432386P.  
XX PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
XX PI Rider TH;  
XX DR WPI; 2003-731477/69.  
XX XX Treating or preventing a pathogen infection in a cell by administering to  
PT the cell chimeric molecules having a pathogen-detection domain or  
PT pathogen-induced product-detection domain and at least one effector  
PT domain.  
XX PS Example 7; SEQ ID NO 272; 348pp; English.  
XX CC The present invention describes a method for treating or preventing a  
CC pathogen infection in a cell. The method comprises administering to the  
CC cell chimeric molecules (I) having at least one pathogen-detection domain  
CC or pathogen-induced product-detection domain and at least one effector  
CC domain. The pathogen-detection domain or pathogen-induced product-  
CC detection domain is not naturally bound to the effector domain. In the  
CC presence of a pathogen or pathogen-induced product in the cell, the  
CC chimeric molecules bind to the pathogen and activate the spread of a pathogen  
CC infection in an organism; (2) a chimeric molecule (I) having at least one  
CC pathogen-detection domain or pathogen-induced product-detection domain  
CC and at least one effector domain; (3) an agent having at least one  
CC pathogen-interacting or pathogen-induced product-interacting molecular  
CC structure and at least one effector-mediated product-interacting molecular  
CC (4) an assay for detecting a pathogen infection in a cell or organism.  
CC (I) has antimicrobial activity, and can be used in gene therapy. The  
CC method is useful for treating or preventing a pathogen infection in a  
CC cell or organism. The present sequence is used in the exemplification of  
XX the present invention.  
XX SQ Sequence 7 AA;  
Query Match 100.0%; Score 27; DB 7; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GRRGR 5  
Db |||||  
1 GRRGR 5  
RESULT 15  
ADG44487  
ID ADG44487 standard; peptide; 7 AA.  
XX AC ADG44487;  
XX DT 26-FEB-2004 (first entry)  
XX DE Anti-pathogen related amino acid sequence SEQ ID NO:286.  
XX KW pathogen infection; chimeric molecule; pathogen-detection domain;  
KW pathogen-induced product-detection domain; effector domain;  
KW antimicrobial; gene therapy.  
XX OS Synthetic.  
XX PN WO2003066003-A2.  
XX PD 14-AUG-2003.  
XX PF 07-FEB-2003; 2003WO-US003978.

```

XX 10-DEC-2002; 2002US-0432386P.
PR 07-FEB-2002; 2002US-0355022P.
XX 07-FEB-2002; 2002US-0355359P.
PR 10-DEC-2002; 2002US-0432386P.
XX 10-DEC-2002; 2002US-0432386P.
XX (NASI ) MASSACHUSETTS INST TECHNOLOGY.
PA Rider TH;
XX WPI; 2003-731477/69.
PI
XX
XX Treating or preventing a pathogen infection in a cell by administering to
DR the cell chimeric molecules having a pathogen-detection domain or
XX pathogen-induced product-detection domain and at least one effector
XX domain.
XX
XX Example 7; SEQ ID NO 286; 348pp; English.
PS
XX The present invention describes a method for treating or preventing a
CC pathogen infection in a cell. The method comprises administering to the
CC cell chimeric molecules (1) having at least one pathogen-detection domain
CC or pathogen-induced product-detection domain and at least one effector
CC domain. The pathogen-detection domain or pathogen-induced product-
CC detection domain is not naturally bound to the effector domain. In the
CC chimeric molecules bind to the pathogen and activate the effector domain.
CC Also described: (1) treating or preventing the spread of a pathogen
CC infection in an organism; (2) a chimeric molecule (1) having at least one
CC pathogen-detection domain or pathogen-induced product-detection domain
CC and at least one effector domain; (3) an agent having at least one
CC pathogen-detection domain or pathogen-induced product-detection domain
CC and at least one effector domain; (3) an agent having at least one
CC pathogen-interacting or pathogen-mediated product-interacting molecular
CC structure and at least one effector-mediated molecular structure; and
CC (4) an assay for detecting a pathogen infection in a cell or organism.
CC (1) has antimicrobial activity, and can be used in gene therapy. The
CC method is useful for treating or preventing a pathogen infection in a
CC cell or organism. The present sequence is used in the exemplification of
CC the present invention.
XX
XX Sequence 7 AA;
SQ
Query Match 100.0%; Score 27; DB 7; Length 7;
Best Local Similarity 100.0%; Pred. No. 2e+06;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5
Db |||||
1 GRGR 5

RESULT 16
ADG44613
ID ADG44613 standard; peptide; 7 AA.
XX
XX ADG44613;
AC
XX
XX 26-FEB-2004 (first entry)
DT
XX
XX Anti-pathogen related amino acid sequence SEQ ID NO:412.
DE
XX pathogen infection; chimeric molecule; pathogen-detection domain;
KW pathogen-induced product-detection domain; effector domain;
KW antimicrobial; gene therapy.
XX
XX Synthetic.
OS
XX WO2003066003-A2.
XX
XX 14-AUG-2003.
PN
XX
XX 07-FEB-2003; 2003WO-US003978.
PD
XX
XX 07-FEB-2002; 2002US-0355022P.
PF
XX 07-FEB-2002; 2002US-0355359P.
XX
XX 07-FEB-2002; 2002US-0355022P.
PR
XX 07-FEB-2002; 2002US-0355359P.
PR
XX (NASI ) MASSACHUSETTS INST TECHNOLOGY.
PA

```

```

PR 10-DEC-2002; 2002US-0432386P.
XX (NASI ) MASSACHUSETTS INST TECHNOLOGY.
PA Rider TH;
XX WPI; 2003-731477/69.
DR
XX
XX Treating or preventing a pathogen infection in a cell by administering to
PT the cell chimeric molecules having a pathogen-detection domain or
PT pathogen-induced product-detection domain and at least one effector
PT domain.
XX
XX Example 7; SEQ ID NO 412; 348pp; English.
PS
XX The present invention describes a method for treating or preventing a
CC pathogen infection in a cell. The method comprises administering to the
CC cell chimeric molecules (1) having at least one pathogen-detection domain
CC or pathogen-induced product-detection domain and at least one effector
CC domain. The pathogen-detection domain or pathogen-induced product-
CC detection domain is not naturally bound to the effector domain. In the
CC presence of a pathogen or pathogen-induced product in the cell, the
CC chimeric molecules bind to the pathogen and activate the effector domain.
CC Also described: (1) treating or preventing the spread of a pathogen
CC infection in an organism; (2) a chimeric molecule (1) having at least one
CC pathogen-detection domain or pathogen-induced product-detection domain
CC and at least one effector domain; (3) an agent having at least one
CC pathogen-interacting or pathogen-mediated product-interacting molecular
CC structure and at least one effector-mediated molecular structure; and
CC (4) an assay for detecting a pathogen infection in a cell or organism.
CC (1) has antimicrobial activity, and can be used in gene therapy. The
CC method is useful for treating or preventing a pathogen infection in a
CC cell or organism. The present sequence is used in the exemplification of
CC the present invention.
XX
XX Sequence 7 AA;
SQ
Query Match 100.0%; Score 27; DB 7; Length 7;
Best Local Similarity 100.0%; Pred. No. 2e+06;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5
Db |||||
1 GRGR 5

RESULT 17
ADG44501
ID ADG44501 standard; peptide; 7 AA.
XX
XX ADG44501;
AC
XX
XX 26-FEB-2004 (first entry)
DT
XX
XX Anti-pathogen related amino acid sequence SEQ ID NO:300.
DE
XX pathogen infection; chimeric molecule; pathogen-detection domain;
KW pathogen-induced product-detection domain; effector domain;
KW antimicrobial; gene therapy.
XX
XX Synthetic.
OS
XX WO2003066003-A2.
XX
XX 14-AUG-2003.
PN
XX
XX 07-FEB-2003; 2003WO-US003978.
PD
XX
XX 07-FEB-2002; 2002US-0355022P.
PR
XX 07-FEB-2002; 2002US-0355359P.
XX
XX 10-DEC-2002; 2002US-0432386P.
XX
XX (NASI ) MASSACHUSETTS INST TECHNOLOGY.
PA

```

```

XX      Rider TH;
XX      WPI; 2003-731477/69.
XX      Treating or preventing a pathogen infection in a cell by administering to
XX      the cell chimeric molecules having a pathogen-detection domain or
XX      pathogen-induced product-detection domain and at least one effector
XX      domain.
XX      Example 7; SEQ ID NO 300; 348pp; English.
XX      The present invention describes a method for treating or preventing a
XX      pathogen infection in a cell. The method comprises administering to the
XX      cell chimeric molecules (I) having at least one pathogen-detection domain
XX      or pathogen-induced product-detection domain and at least one effector
XX      domain. The pathogen-detection domain or pathogen-induced product-
XX      detection domain is not naturally bound to the effector domain. In the
XX      presence of a pathogen or pathogen-induced product in the cell, the
XX      chimeric molecules bind to the pathogen and activate the effector domain.
XX      Also described: (1) treating or preventing the spread of a pathogen
XX      infection in an organism; (2) a chimeric molecule (I) having at least one
XX      pathogen-detection domain or pathogen-induced product-detection domain
XX      and at least one effector domain; (3) an agent having at least one
XX      pathogen-interacting or pathogen-induced product-interacting molecular
XX      structure and at least one effector domain; (4) an agent having at least one
XX      pathogen-interacting or pathogen-induced product-interacting molecular
XX      structure and at least one effector-mediated product-interacting molecular
XX      structure; and (4) an assay for detecting a pathogen infection in a cell or organism.
XX      (I) has antimicrobial activity, and can be used in gene therapy. The
XX      method is useful for treating or preventing a pathogen infection in a
XX      cell or organism. The present sequence is used in the exemplification of
XX      the present invention.
XX      Sequence 7 AA;
XX      Query Match      100.0%; Score 27; DB 7; Length 7;
XX      Best Local Similarity 100.0%; Pred. No. 2e+06;
XX      Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX      QY      1 GRRGR 5
XX      Db      |||||
XX      1 GRRGR 5
XX      RESULT 18
XX      ADG44599
XX      ID      ADG44599 standard; peptide; 7 AA.
XX      AC      ADG44599;
XX      DT      26-FEB-2004 (first entry)
XX      DE      Anti-pathogen related amino acid sequence SEQ ID NO:398.
XX      KW      pathogen infection; chimeric molecule; pathogen-detection domain;
XX      KW      pathogen-induced product-detection domain; effector domain;
XX      KW      antimicrobial; gene therapy.
XX      OS      Synthetic.
XX      PN      WO2003066003-A2.
XX      PD      14-AUG-2003.
XX      PF      07-FEB-2003; 2003WO-US003978.
XX      PR      07-FEB-2002; 2002US-0355022P.
XX      PR      07-FEB-2002; 2002US-0355359P.
XX      PR      10-DEC-2002; 2002US-0432386P.
XX      PA      (MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX      PI      Rider TH;
XX      DR      WPI; 2003-731477/69.
XX      Treating or preventing a pathogen infection in a cell by administering to

```

```

DR      WPI; 2003-731477/69.
XX      Treating or preventing a pathogen infection in a cell by administering to
XX      the cell chimeric molecules having a pathogen-detection domain or
XX      pathogen-induced product-detection domain and at least one effector
XX      domain.
XX      Example 7; SEQ ID NO 398; 348pp; English.
XX      The present invention describes a method for treating or preventing a
XX      pathogen infection in a cell. The method comprises administering to the
XX      cell chimeric molecules (I) having at least one pathogen-detection domain
XX      or pathogen-induced product-detection domain and at least one effector
XX      domain. The pathogen-detection domain or pathogen-induced product-
XX      detection domain is not naturally bound to the effector domain. In the
XX      presence of a pathogen or pathogen-induced product in the cell, the
XX      chimeric molecules bind to the pathogen and activate the effector domain.
XX      Also described: (1) treating or preventing the spread of a pathogen
XX      infection in an organism; (2) a chimeric molecule (I) having at least one
XX      pathogen-detection domain or pathogen-induced product-detection domain
XX      and at least one effector domain; (3) an agent having at least one
XX      pathogen-interacting or pathogen-induced product-interacting molecular
XX      structure and at least one effector-mediated product-interacting molecular
XX      structure; and (4) an assay for detecting a pathogen infection in a cell or organism.
XX      (I) has antimicrobial activity, and can be used in gene therapy. The
XX      method is useful for treating or preventing a pathogen infection in a
XX      cell or organism. The present sequence is used in the exemplification of
XX      the present invention.
XX      Sequence 7 AA;
XX      Query Match      100.0%; Score 27; DB 7; Length 7;
XX      Best Local Similarity 100.0%; Pred. No. 2e+06;
XX      Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX      QY      1 GRRGR 5
XX      Db      |||||
XX      1 GRRGR 5
XX      RESULT 19
XX      ADG44585
XX      ID      ADG44585 standard; peptide; 7 AA.
XX      AC      ADG44585;
XX      DT      26-FEB-2004 (first entry)
XX      DE      Anti-pathogen related amino acid sequence SEQ ID NO:384.
XX      KW      pathogen infection; chimeric molecule; pathogen-detection domain;
XX      KW      pathogen-induced product-detection domain; effector domain;
XX      KW      antimicrobial; gene therapy.
XX      OS      Synthetic.
XX      PN      WO2003066003-A2.
XX      PD      14-AUG-2003.
XX      PF      07-FEB-2003; 2003WO-US003978.
XX      PR      07-FEB-2002; 2002US-0355022P.
XX      PR      07-FEB-2002; 2002US-0355359P.
XX      PR      10-DEC-2002; 2002US-0432386P.
XX      PA      (MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX      PI      Rider TH;
XX      DR      WPI; 2003-731477/69.
XX      Treating or preventing a pathogen infection in a cell by administering to

```

PT the cell chimeric molecules having a pathogen-detection domain or  
PT pathogen-induced product-detection domain and at least one effector  
PT domain.  
XX  
PS Example 7; SEQ ID NO 384; 348pp; English.  
XX  
CC The present invention describes a method for treating or preventing a  
CC pathogen infection in a cell. The method comprises administering to the  
CC cell chimeric molecules (1) having at least one pathogen-detection domain  
CC or pathogen-induced product-detection domain and at least one effector  
CC domain. The pathogen-detection domain or pathogen-induced product-  
CC detection domain is not naturally bound to the effector domain. In the  
CC presence of a pathogen or pathogen-induced product in the cell, the  
CC chimeric molecules bind to the pathogen and activate the spread of a pathogen  
CC Also described: (1) treating or preventing the spread of a pathogen  
CC infection in an organism; (2) a chimeric molecule (1) having at least one  
CC pathogen-detection domain or pathogen-induced product-detection domain  
CC and at least one effector domain; (3) an agent having at least one  
CC pathogen-interacting or pathogen-induced product-interacting molecular  
CC structure and at least one effector-mediated molecular structure; and  
CC (4) an assay for detecting a pathogen infection in a cell or organism.  
CC (1) has antimicrobial activity, and can be used in gene therapy. The  
CC method is useful for treating or preventing a pathogen infection in a  
CC cell or organism. The present sequence is used in the exemplification of  
CC the present invention.  
XX  
SQ Sequence 7 AA;  
  
Query Match 100.0%; Score 27; DB 7; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GRRGR 5  
DB |||||  
1 GRRGR 5  
  
RESULT 20  
ADG44529  
ID ADG44529 standard; peptide; 7 AA.  
XX  
AC ADG44529;  
XX  
DT 26-FEB-2004 (first entry)  
XX  
DE Anti-pathogen related amino acid sequence SEQ ID NO:328.  
XX  
KW pathogen infection; chimeric molecule; pathogen-detection domain;  
KW pathogen-induced product-detection domain; effector domain;  
KW antimicrobial; gene therapy.  
XX  
OS Synthetic.  
XX  
PN WO2003066003-A2.  
XX  
PD 14-AUG-2003.  
XX  
PF 07-FEB-2003; 2003WO-US003978.  
XX  
PR 07-FEB-2002; 2002US-0355022P.  
XX  
PR 07-FEB-2002; 2002US-0355359P.  
XX  
PR 10-DEC-2002; 2002US-0432386P.  
XX  
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
XX  
PI Rider TH;  
XX  
DR WPI; 2003-731477/69.  
XX  
PT Treating or preventing a pathogen infection in a cell by administering to  
PT the cell chimeric molecules having a pathogen-detection domain or  
PT pathogen-induced product-detection domain and at least one effector  
PT domain.

XX  
PS Example 7; SEQ ID NO 328; 348pp; English.  
XX  
CC The present invention describes a method for treating or preventing a  
CC pathogen infection in a cell. The method comprises administering to the  
CC cell chimeric molecules (1) having at least one pathogen-detection domain  
CC or pathogen-induced product-detection domain and at least one effector  
CC domain. The pathogen-detection domain or pathogen-induced product-  
CC detection domain is not naturally bound to the effector domain. In the  
CC presence of a pathogen or pathogen-induced product in the cell, the  
CC chimeric molecules bind to the pathogen and activate the spread of a pathogen  
CC Also described: (1) treating or preventing the spread of a pathogen  
CC infection in an organism; (2) a chimeric molecule (1) having at least one  
CC pathogen-detection domain or pathogen-induced product-detection domain  
CC and at least one effector domain; (3) an agent having at least one  
CC pathogen-interacting or pathogen-induced product-interacting molecular  
CC structure and at least one effector-mediated molecular structure; and  
CC (4) an assay for detecting a pathogen infection in a cell or organism.  
CC (1) has antimicrobial activity, and can be used in gene therapy. The  
CC method is useful for treating or preventing a pathogen infection in a  
CC cell or organism. The present sequence is used in the exemplification of  
CC the present invention.  
XX  
SQ Sequence 7 AA;  
  
Query Match 100.0%; Score 27; DB 7; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GRRGR 5  
DB |||||  
1 GRRGR 5  
  
RESULT 21  
ADG44627  
ID ADG44627 standard; peptide; 7 AA.  
XX  
AC ADG44627;  
XX  
DT 26-FEB-2004 (first entry)  
XX  
DE Anti-pathogen related amino acid sequence SEQ ID NO:426.  
XX  
KW pathogen infection; chimeric molecule; pathogen-detection domain;  
KW pathogen-induced product-detection domain; effector domain;  
KW antimicrobial; gene therapy.  
XX  
OS Synthetic.  
XX  
PN WO2003066003-A2.  
XX  
PD 14-AUG-2003.  
XX  
PF 07-FEB-2003; 2003WO-US003978.  
XX  
PR 07-FEB-2002; 2002US-0355022P.  
XX  
PR 07-FEB-2002; 2002US-0355359P.  
XX  
PR 10-DEC-2002; 2002US-0432386P.  
XX  
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
XX  
PI Rider TH;  
XX  
DR WPI; 2003-731477/69.  
XX  
PT Treating or preventing a pathogen infection in a cell by administering to  
PT the cell chimeric molecules having a pathogen-detection domain or  
PT pathogen-induced product-detection domain and at least one effector  
PT domain.  
XX  
PS Example 7; SEQ ID NO 426; 348pp; English.  
XX

CC The present invention describes a method for treating or preventing a  
CC pathogen infection in a cell. The method comprises administering to the  
CC cell chimeric molecules (I) having at least one pathogen-detection domain  
CC or pathogen-induced product-detection domain and at least one effector  
CC domain. The pathogen-detection domain or pathogen-induced product-  
CC detection domain is not naturally bound to the effector domain. In the  
CC presence of a pathogen or pathogen-induced product in the cell, the  
CC chimeric molecules bind to the pathogen and activate the effector domain.  
CC Also described: (1) treating or preventing the spread of a pathogen  
CC infection in an organism; (2) a chimeric molecule (I) having at least one  
CC pathogen-detection domain or pathogen-induced product-detection domain  
CC and at least one effector domain; (3) an agent having at least one  
CC pathogen-interacting or pathogen-induced product-interacting molecular  
CC structure and at least one effector-mediated product-molecular structure; and  
CC (4) an assay for detecting a pathogen infection in a cell or organism.  
CC (I) has antimicrobial activity, and can be used in gene therapy. The  
CC method is useful for treating or preventing a pathogen infection in a  
CC cell or organism. The present sequence is used in the exemplification of  
CC the present invention.

XX Sequence 7 AA;  
SQ Query Match 100.0%; Score 27; DB 7; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
Db |||||  
1 GRRGR 5

RESULT 22  
AAW86181  
ID AAW86181 standard; peptide; 8 AA.  
AC AAW86181;  
XX  
DT 04-MAR-1999 (first entry)  
XX  
DE Peptide used in a method for inhibiting wound contraction.  
KW Wound contraction; reduction; inhibition; tissue regeneration; scar;  
KW wound; joint motion; body deformation.  
XX Synthetic.  
OS  
XX US5851994-A.  
XX  
PD 22-DEC-1998.  
XX  
PF 06-JUN-1995; 95US-00473025.  
XX  
PR 28-APR-1994; 94US-00234979.  
PA (LJOL-) LA JOLLA CANCER RES FOUND.  
XX  
PI Polarek J, Schreiber R;  
XX  
DR WPI; 1999-080478/07.  
XX  
PT Inhibition of wound contraction - with peptide derivatives rich in basic  
PT amino acids.  
XX  
PS Claim 11; Col 24; 16pp; English.

CC The invention provides methods for reduction or inhibition of wound  
CC contraction that comprises administration of a peptide having more than 3  
CC consecutive basic amino acid residues. Alternatively, the peptide  
CC contains the amino acid sequence Arg-Gly-Asp and a basic amino acid  
CC sequence, or the peptide comprises 6-30 amino acids in which at least 4  
CC out of a sequence of 6 consecutive amino acids are basic amino acids. The  
CC method is used to allow normal tissue regeneration without excessive scar  
CC formation which, in the case of large wounds, can result in loss of joint

CC motion or major body deformation. Sequences AAW86170 to AAW86183  
CC represent specifically claimed examples of peptides that can be used in  
CC the method of the invention

XX Sequence 8 AA;  
SQ Query Match 100.0%; Score 27; DB 2; Length 8;  
Best Local Similarity 100.0%; Pred. No. 2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
Db |||||  
1 GRRGR 5

RESULT 23  
ABP23953  
ID ABP23953 standard; peptide; 8 AA.  
XX  
AC ABP23953;  
XX  
DT 11-SEP-2003 (revised)  
DT 15-JUL-2002 (first entry)  
XX  
DE HIV All motif vpr peptide #1.  
XX  
KW HIV; HIV-1; human immunodeficiency virus; env; pol; gag; nef; vpr; vpu;  
KW vif; tat; cytotoxic T lymphocyte; CTL; immune response; epitope; antigen;  
KW vaccine; HIV infection; immunisation; virucide.  
XX  
OS Human immunodeficiency virus 1.  
XX  
FN WO200124810-A1.  
XX  
PD 12-APR-2001.  
XX  
PF 05-OCT-2000; 2000WO-US027766.  
XX  
PR 05-OCT-1999; 99US-00412863.  
XX (EPIM-) EPIMUNE INC.  
PA Sette A, Sidney J, Southwood S, Livingston BD, Chesnut R;  
PI Baker DM, Cellis E, Kubo RT, Grey HM;  
XX  
DR WPI; 2001-354887/37.  
XX  
PT Vaccine compositions comprising human immunodeficiency virus-1 (HIV-1)  
PT peptide groups, useful for vaccinating against HIV-1.  
XX  
PS Claim 32; Page 359; 448pp; English.

CC The present invention describes a composition (I) comprising a prepared  
CC human immunodeficiency virus-1 (HIV-1) group comprising an amino acid  
CC sequence selected from 51 defined amino acid sequences (ABL25347 to  
CC ABP25397). (I) has virucide activity and can be used in vaccines. (I) may  
CC be used for immunising subjects against HIV-1 infections. The use of  
CC group-based vaccines has several advantages over traditional vaccines,  
CC particularly when compared to the use of whole antigens in vaccine  
CC compositions. There is evidence that the immune response to whole  
CC antigens is directed largely toward variable regions of the antigen,  
CC allowing for immune escape due to mutations. The groups for inclusion in  
CC an group-based vaccine may be selected from conserved regions of viral or  
CC tumour-associated antigens, which therefore reduces the likelihood of  
CC escape mutants. Furthermore, immunosuppressive groups that may be present  
CC in whole antigens can be avoided with the use of group-based vaccines. An  
CC additional advantage of an group-based vaccine approach is the ability to  
CC combine selected groups (CTL and HTL), and further, to modify the  
CC composition of the groups, achieving, for example, enhanced  
CC immunogenicity. Accordingly, the immune response can be modulated, as  
CC appropriate, for the target disease. Similar engineering of the response  
CC is not possible with traditional approaches. ABP11501 to ABP25412  
CC represent peptide sequences used in the exemplification of the present

```

CC invention. (Updated on 11-SEP-2003 to standardise OS field)
XX
SQ Sequence 8 AA;

Query Match      100.0%; Score 27; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 2e+06;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5
Db 4 GRRGR 8
    |||||

RESULT 24
ABG31118
ID ABG31118 standard; peptide; 8 AA.
XX
AC ABG31118;
XX
DT 21-OCT-2002 (first entry)
XX
DE Cathepsin-B sensitive chromophore attachment spacer #4.
XX
KW Chromophore; imaging probe; spacer; cathepsin-B; inflammation;
KW rheumatoid arthritis; cancer; cardiovascular disease; atherosclerosis;
KW dermatological disease; Kaposi's sarcoma; psoriasis; ophthalmic disease;
KW diabetic retinopathy; infectious disease; immunological disease;
KW acquired immunodeficiency syndrome; AIDS; neurodegenerative disease;
KW Alzheimer's disease; bone-related disease; osteoporosis;
KW environmental disease.
XX
OS Synthetic.
XX
FH Key      Location/Qualifiers
FT Modified-site 7
FT /label= OTHER
FT /note= "Lys is covalently attached to an FITC
FT (fluorescein isothiocyanate) moiety"
FT Modified-site 8
FT /label= OTHER
FT /note= "Cys is amidated and covalently linked to a QSY7
FT (not defined) moiety"
XX
PN WO200256570-A2.
XX
XX 25-JUL-2002.
XX
XX 07-JAN-2002; 2002WO-US000379.
XX
XX 05-JAN-2001; 2001US-0260123P.
XX 19-MAR-2001; 2001US-0277352P.
XX 09-NOV-2001; 2001US-00277352.
XX
XX (GEO ) GEN HOSPITAL CORP.
XX
XX Weissleder R, Tung C, Mahmood U;
XX
XX WPI; 2002-590684/63.
XX
XX Activatable imaging probe for in vivo target optical imaging, has
XX chromophore attachment moiety chemically linked to chromophores, so that
XX upon activation of probe, optical properties of chromophores are altered.
XX
XX Disclosure; Page 11; 69pp; English.
XX
XX The invention relates to an activatable imaging probe (I) (activated by
XX phosphorylation, dephosphorylation, pH mediated cleavage, conformation
XX change, enzyme-mediated splicing, enzyme-mediated transfer of the one or
XX more chromophores, hybridisation of a nucleic acid sequence to a
XX complementary target nucleic acid, binding of the probe to an analyte,
XX chemical modification of the chromophore or binding of the probe to a
XX receptor) comprises a chromophore attachment moiety (II) and one or more
XX chromophores (III), where (III) is chemically linked to (II), so that

```

```

CC upon activation of (I), the optical properties of (III) are altered. The
CC probe is useful for in vivo optical imaging of a target in a subject, by
CC delivering the probe to the subject, allowing adequate time for
CC activation of the probe within the target, illuminating the target with
CC light of a wavelength absorbable by the chromophores, detecting a signal
CC emitted by the chromophores, optionally, repeating the steps at
CC predetermined intervals to enable evaluation of the emitted signal of the
CC chromophores in the subject over time, and forming an optical image from
CC the emitted signal. The method is useful for detecting a disease e.g.
CC inflammation (e.g. rheumatoid arthritis) cancer, cardiovascular disease
CC (e.g. atherosclerosis), dermatological disease (e.g. Kaposi's sarcoma and
CC psoriasis), ophthalmic disease (e.g. diabetic retinopathy), infectious
CC disease, immunological diseases (e.g. acquired immunodeficiency syndrome,
CC AIDS), neurodegenerative disease (e.g. Alzheimer's disease and bone-
CC related disease (e.g. osteoporosis) in the subject, for characterising a
CC phenotype or genotype of a disease in the subject and for characterising
CC the severity of a disease. The probe is also useful in in vivo imaging
CC for simultaneous imaging of one or more different targets in a subject,
CC in an optical imaging method for assessing activity of an agent in a
CC subject, by carrying out the above said method, administering the agent
CC to the subjects and repeating the above said steps, and comparing the
CC emitted signals and images over time or at a different agent dose to
CC assess the activity of the agent. The probe is useful for determining the
CC presence of a composition (e.g. drug or a polypeptide expressed by a
CC gene) in a subject. The illumination and detecting processes are carried
CC out using endoscope, catheter, tomographic system, surgical goggles with
CC attached bandpass filters or an intraoperative microscope. The probe is
CC useful for assessing the effective dosage of an agent in a subject, in an
CC optical imaging method for guiding therapeutic interventions (e.g.
CC surgical interventions) in a subject such as mammal e.g. human or animal
CC model of a disease, and in in vitro optical imaging method for assessing
CC the activity of an agent in a sample. The present sequence is a peptide
CC spacer containing a chromophore activation site
XX
XX Sequence 8 AA;

Query Match      100.0%; Score 27; DB 5; Length 8;
Best Local Similarity 100.0%; Pred. No. 2e+06;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5
Db 1 GRRGR 5
    |||||

RESULT 25
ADD68832
ID ADD68832 standard; peptide; 8 AA.
XX
XX AC ADD68832;
XX
XX 15-JAN-2004 (first entry)
XX
XX Cathepsin D specific substrate peptide - SEQ ID 7.
XX
XX optical imaging probe; chromophore; cancer; cardiovascular disease;
XX neurodegenerative; immunologic; autoimmune; inherited; infectious; bone;
XX environmental; cathepsin D specific substrate.
XX
XX Unidentified.
XX
XX Key      Location/Qualifiers
XX Modified-site 7
XX /label= OTHER
XX /note= "OTHER = Labelled with FITC"
XX Modified-site 8
XX /label= OTHER
XX /note= "OTHER = C-terminal amide"
XX
XX US2003044353-A1.
XX
XX 06-MAR-2003.
XX
XX

```

[illegible]



XX New nucleic acid molecule encoding nucleocapsid- or spike-gene protein of  
PT a human Severe Acute Respiratory Syndrome (hSARS) virus, useful for  
PT diagnosing and treating SARS.  
XX  
XX Example; SEQ ID NO 806; 200pp; English.  
XX  
XX The present invention relates to novel human Severe Acute Respiratory  
CC Syndrome (hSARS) viral nucleic acid and protein sequences derived from a  
CC hSARS virus having China Center for Type Culture Collection Deposit  
CC Accession No. CCTCC-V200303. The present invention also relates to novel  
CC nucleic acid molecules (1; ADT41483 or ADT41485) encoding a nucleocapsid-  
CC (N) or spike (S)-gene protein of a hSARS virus. Also disclosed are  
CC methods for detecting the presence of a N- or S-gene of the hSARS virus  
CC or of the protein in a biological sample and identifying a subject  
CC infected with the hSARS virus. The hSARS virus, nucleic acid and protein  
CC sequences are useful as vaccines for diagnosing or treating SARS. They  
CC are also useful in clinical and scientific research applications. The  
CC hSARS virus genome (ADT39027) was obtained and the amino acid sequences  
CC of all three reading frames were deduced. ADT39028 is the full-length  
CC protein encoded by the first reading frame and ADT39029-ADT39251 are the  
CC peptides from the first reading frame protein. ADT39252 is the full-  
CC length protein encoded by the second reading frame and ADT39253-ADT39748  
CC are the peptides from the second reading frame protein. ADT39749 is the  
CC full-length protein encoded by the third reading frame and ADT39750-  
CC ADT40119 are the peptides from the third reading frame protein.  
XX  
XX Sequence 8 AA;

Query Match 100.0%; Score 27; DB 8; Length 8;  
Best Local Similarity 100.0%; Pred. No. 2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRRGR 5  
|||||  
Db 1 GRRGR 5

## RESULT 28

ADT39236  
ID ADS79236 standard; protein; 8 AA.

XX AC ADS79236;

XX 30-DEC-2004 (first entry)

XX SARS virus reading frame 3 protein #69.

XX virucide; vaccine; detection; severe acute respiratory syndrome;  
KW real-time quantitative polymerase chain reaction; SARS.

XX SARS coronavirus.

XX WO2004085455-A1.

XX 07-OCT-2004.

XX 24-MAR-2004; 2004WO-CN000247.

XX 24-MAR-2003; 2003US-0457031P.

XX 26-MAR-2003; 2003US-0457730P.

XX 02-APR-2003; 2003US-0459931P.

XX 03-APR-2003; 2003US-0460357P.

XX 08-APR-2003; 2003US-0461265P.

XX 14-APR-2003; 2003US-0462805P.

XX 23-APR-2003; 2003US-0468139P.

XX 05-MAY-2003; 2003US-0471200P.

XX (UYHK-) UNIV HONG KONG.

XX Chan K, Guan Y, Nicholls JM, Peiris JSM, Poon L, Yuen K;  
XX  
XX

DR WPI; 2004-737292/72.  
DR N-PSDB; ADS78448.  
XX  
XX New isolated nucleic acid molecule useful for detecting, treating,  
PT ameliorating, or preventing the virus causing severe acute respiratory  
PT syndrome in humans using a real-time quantitative polymerase chain  
PT reaction assay.  
XX  
XX Example; SEQ ID NO 806; 183pp; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule consisting  
CC essentially of, and/or hybridizes under stringent conditions to a fully  
CC defined nucleotide sequence of 16-25 base pairs (bp; SEQ ID NO: 2471-  
CC 2476), or its complement. The methods and compositions of the present  
CC invention are useful for the detection of the virus causing Severe Acute  
CC Respiratory Syndrome (SARS) in humans using a real-time quantitative  
CC polymerase chain reaction (PCR) assay. They can also be used in treating,  
CC ameliorating, managing or preventing SARS. This sequence corresponds to a  
CC partial SARS protein sequence from reading frame 3.  
XX  
XX Sequence 8 AA;

Query Match 100.0%; Score 27; DB 8; Length 8;  
Best Local Similarity 100.0%; Pred. No. 2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRRGR 5  
|||||  
Db 1 GRRGR 5

## RESULT 29

ADT37348

ID ADT37348 standard; peptide; 8 AA.

XX AC ADT37348;

XX 30-DEC-2004 (first entry)

XX hsSARS virus peptide, SEQ ID 806.

XX Virucide; Severe Acute Respiratory Syndrome; SARS; vaccine.

XX SARS coronavirus.

XX WO2004085633-A1.

XX 07-OCT-2004.

XX 24-MAR-2004; 2004WO-CN000248.

XX 24-MAR-2003; 2003US-0457031P.

XX 26-MAR-2003; 2003US-0457730P.

XX 02-APR-2003; 2003US-0459931P.

XX 03-APR-2003; 2003US-0460357P.

XX 08-APR-2003; 2003US-0461265P.

XX 14-APR-2003; 2003US-0462805P.

XX 23-APR-2003; 2003US-0464886P.

XX (UYHK-) UNIV HONG KONG.

XX Chan K, Guan Y, Nicholls JM, Peiris JSM, Poon L, Yuen K;  
XX Leung FC;

XX WPI; 2004-728736/71.

XX New isolated human severe acute respiratory syndrome (hSARS) virus,  
PT useful as vaccine for diagnosing or treating SARS or in clinical and  
PT scientific research applications.  
XX  
XX Example; SEQ ID NO 806; 176pp; English.

XX The present invention relates to novel human Severe Acute Respiratory

CC Syndrome (hSARS) viral nucleic acid and protein sequences derived from a  
CC hSARS virus having China Center for Type Culture Collection Deposit  
CC Accession No. CCTCC-V200303. The hSARS virus, nucleic acid and protein  
CC sequences are useful as vaccines for diagnosing or treating SARS. They  
CC are also useful in clinical and scientific research applications. The  
CC hSARS virus genome (ADT36557) was obtained and the amino acid sequences  
CC of all three reading frames were deduced. ADT36558 is the full-length  
CC protein encoded by the first reading frame and ADT36559-ADT36781 are the  
CC peptides from the first reading frame protein. ADT36782 is the full-  
CC length protein encoded by the second reading frame and ADT36783-ADT37278  
CC are the peptides from the second reading frame protein. ADT37279 is the  
CC full-length protein encoded by the third reading frame and ADT37280-  
CC ADT37649 are the peptides from the third reading frame protein.  
XX  
SQ Sequence 8 AA;  
Query Match 100.0%; Score 27; DB 8; Length 8;  
Best Local Similarity 100.0%; Pred. No. 2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GRRGR 5  
Db |||||  
1 GRRGR 5  
RESULT 30  
ID ABY00351  
XX ABY00351 standard; peptide; 8 AA.  
AC ABY00351;  
XX  
DT 16-JUN-2005 (first entry)  
XX  
DE SARS coronavirus genome reading frame +3-encoded peptide, SEQ:6878.  
XX  
KW Vaccine; nucleic acid vaccine; drug screening; diagnosis;  
KW SARS coronavirus infection; infection; respiratory disease; virucide.  
XX  
OS SARS coronavirus.  
XX  
PN WO2004092360-A2.  
XX  
PD 28-OCT-2004.  
XX  
PF 09-APR-2004; 2004WO-US011710.  
XX  
PR 10-APR-2003; 2003US-0462218P.  
XX  
PR 11-APR-2003; 2003US-0462465P.  
PR 12-APR-2003; 2003US-0462418P.  
PR 13-APR-2003; 2003US-0462748P.  
PR 14-APR-2003; 2003US-0463109P.  
PR 15-APR-2003; 2003US-0463460P.  
PR 16-APR-2003; 2003US-0463668P.  
PR 17-APR-2003; 2003US-0463983P.  
PR 18-APR-2003; 2003US-0463971P.  
PR 22-APR-2003; 2003US-0464838P.  
PR 22-APR-2003; 2003US-0464899P.  
PR 23-APR-2003; 2003US-0465273P.  
PR 24-APR-2003; 2003US-0465335P.  
PR 05-MAY-2003; 2003US-0468312P.  
PR 22-MAY-2003; 2003US-0473144P.  
PR 14-AUG-2003; 2003US-0495024P.  
PR 23-SEP-2003; 2003US-0505652P.  
PR 11-OCT-2003; 2003US-0510781P.  
PR 11-DEC-2003; 2003US-0529464P.  
PR 12-JAN-2004; 2004US-0536177P.  
PR 07-APR-2004; 2004US-0560757P.  
XX  
PA (CHIR ) CHIRON CORP.  
XX  
PI Rappuoli R, Masignani V, Stadler K, Gregersen J, Chien D, Han J;  
PI Polo J, Weiner A, Houghton M, Song HC, Seo MY, Donnelly JJ;  
PI Klenk HD, Valiante N;

XX  
DR WPI; 2004-766863/75.  
XX  
PT Novel isolated polypeptide e.g. spike polypeptide, Env polypeptide, of  
PT severe acute respiratory syndrome virus (SARS), useful as vaccine for  
PT SARS.  
XX  
PS Disclosure; SEQ ID NO 6878; 839pp; English.  
XX  
CC The invention relates to isolated polypeptides of the severe acute  
CC respiratory syndrome (SARS) coronavirus. The polypeptides include spike  
CC (S or E2), env (E or SM), membrane (M or E1), hemagglutinin-esterase (HE  
CC or E3), and nucleocapsid (N) polypeptides, and the ORF1a and ORF1ab  
CC (replicase) polypeptides and their proteolytic fragments. The invention  
CC also relates to antibodies which recognise the polypeptides; nucleic  
CC acids encoding the SARS virus polypeptides; primers specific for SARS  
CC virus nucleic acid sequences; kits for amplifying SARS virus target  
CC nucleic acids; a double-stranded RNA molecule 10-30 nucleotides in length  
CC which is able to inactivate the SARS virus in a mammalian cell; an  
CC expression construct for recombinant expression of a SARS virus spike  
CC protein; a viral vector for in vivo delivery of a SARS virus polypeptide-  
CC encoding nucleic acid; and a mammalian cell line stably expressing a SARS  
CC viral antigen. The invention additionally provides a vaccine for the  
CC treatment or prevention of SARS comprising an inactivated SARS virus, a  
CC killed SARS virus, an attenuated SARS virus, a split SARS virus  
CC preparation, or at least one purified SARS virus antigens; methods of  
CC making inactivated SARS virus and vaccines containing it; an alpha-virus  
CC replicon particle comprising one or more SARS viral antigens; and a  
CC vaccine comprising one or more SARS virus antigens and one or more  
CC respiratory virus antigens. The invention further encompasses a method of  
CC identifying a therapeutically active agent by measuring the effect of the  
CC agent on a SARS-related enzyme, and a method of treating a SARS patient  
CC using small molecule viral inhibitors. The SARS virus polypeptides and  
CC nucleic acids can be used in the preparation and manufacture of vaccines  
CC for the treatment or prevention of SARS. The SARS virus polypeptides,  
CC antibodies against them, and SARS virus-specific primers and kits  
CC containing them are useful for diagnosing or identifying the presence of  
CC SARS in a biological sample. The present sequence represents a  
CC polypeptide encoded by a SARS coronavirus polynucleotide. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 8 AA;  
Query Match 100.0%; Score 27; DB 8; Length 8;  
Best Local Similarity 100.0%; Pred. No. 2e+06;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GRRGR 5  
Db |||||  
1 GRRGR 5  
RESULT 31  
ID ABP22161  
XX ABP22161 standard; peptide; 9 AA.  
AC ABP22161;  
XX  
DT 11-SEP-2003 (revised)  
DT 15-JUL-2002 (first entry)  
XX  
DE HIV A03 motif vpr peptide #3.  
XX  
KW HIV; HIV-1; human immunodeficiency virus; env; pol; gag; nef; vpr; vpu;  
KW vif; tat; cytotoxic T lymphocyte; CTL; immune response; epitope; antigen;  
KW vaccine; HIV infection; immunisation; virucide.  
XX  
OS Human immunodeficiency virus 1.  
XX  
PN WO200124810-A1.  
XX



CC be used for immunising subjects against HIV-1 infections. The use of  
CC group-based vaccines has several advantages over traditional vaccines,  
CC particularly when compared to the use of whole antigens in vaccine  
CC compositions. There is evidence that the immune response to whole  
CC antigens is directed largely toward variable regions of the antigen,  
CC allowing for immune escape due to mutations. The groups for inclusion in  
CC an group-based vaccine may be selected from conserved regions of viral or  
CC tumour-associated antigens, which therefore reduces the likelihood of  
CC escape mutants. Furthermore, immunosuppressive groups that may be present  
CC in whole antigens can be avoided with the use of group-based vaccines. An  
CC additional advantage of an group-based vaccine approach is the ability to  
CC combine selected groups (CTL and HTL), and further, to modify the  
CC composition of the groups, achieving, for example, enhanced  
CC immunogenicity. Accordingly, the immune response can be modulated, as  
CC appropriate, for the target disease. Similar engineering of the response  
CC is not possible with traditional approaches. ABP11501 to ABP25412  
CC represent peptide sequences used in the exemplification of the present  
CC invention. (Updated on 11-SEP-2003 to standardise OS field)

XX  
SQ Sequence 10 AA;  
Query Match 100.0%; Score 27; DB 4; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
DB 3 GRRGR 7

RESULT 34  
ADG44611  
ID ADG44611 standard; peptide; 10 AA.

XX  
AC ADG44611;  
DT 26-FEB-2004 (first entry)

XX Anti-pathogen related amino acid sequence SEQ ID NO:410.

XX pathogen infection; chimeric molecule; pathogen-detection domain;  
KW pathogen-induced product-detection domain; effector domain;  
XX antimicrobial; gene therapy.

OS Synthetic.

XX WO2003066003-A2.

XX 14-AUG-2003.

XX 07-FEB-2003; 2003WO-US003978.

XX 07-FEB-2002; 2002US-0355022P.

PR 07-FEB-2002; 2002US-0355359P.

PR 10-DEC-2002; 2002US-0432386P.

XX (MASI ) MASSACHUSETTS INST TECHNOLOGY.

PA Rider TH;

XX WPI; 2003-731477/69.

XX Treating or preventing a pathogen infection in a cell by administering to  
PT the cell chimeric molecules having a pathogen-detection domain or  
PT pathogen-induced product-detection domain and at least one effector  
PT domain.

XX Example 7; SEQ ID NO 410; 348pp; English.

XX The present invention describes a method for treating or preventing a  
CC pathogen infection in a cell. The method comprises administering to the  
CC cell chimeric molecules (I) having at least one pathogen-detection domain  
CC or pathogen-induced product-detection domain and at least one effector

CC domain. The pathogen-detection domain or pathogen-induced product-  
CC detection domain is not naturally bound to the effector domain. In the  
CC presence of a pathogen or pathogen-induced product in the cell, the  
CC chimeric molecules bind to the pathogen and activate the effector domain.  
CC Also described: (1) treating or preventing the spread of a pathogen  
CC infection in an organism; (2) a chimeric molecule (I) having at least one  
CC pathogen-detection domain or pathogen-induced product-detection domain  
CC and at least one effector domain; (3) an agent having at least one  
CC pathogen-interacting or pathogen-induced product-interacting molecular  
CC structure and at least one effector-mediating molecular structure; and  
CC (4) an assay for detecting a pathogen infection in a cell or organism.  
CC (I) has antimicrobial activity, and can be used in gene therapy. The  
CC method is useful for treating or preventing a pathogen infection in a  
CC cell or organism. The present sequence is used in the exemplification of  
CC the present invention.

XX Sequence 10 AA;

Query Match 100.0%; Score 27; DB 7; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
DB 1 GRRGR 5

RESULT 35  
ADG44569  
ID ADG44569 standard; peptide; 10 AA.

XX  
AC ADG44569;

XX 26-FEB-2004 (first entry)

XX Anti-pathogen related amino acid sequence SEQ ID NO:368.

XX pathogen infection; chimeric molecule; pathogen-detection domain;  
KW pathogen-induced product-detection domain; effector domain;  
XX antimicrobial; gene therapy.

OS Synthetic.

XX WO2003066003-A2.

XX 14-AUG-2003.

XX 07-FEB-2003; 2003WO-US003978.

XX 07-FEB-2002; 2002US-0355022P.

PR 07-FEB-2002; 2002US-0355359P.

PR 10-DEC-2002; 2002US-0432386P.

XX (MASI ) MASSACHUSETTS INST TECHNOLOGY.

XX Rider TH;

XX WPI; 2003-731477/69.

XX Treating or preventing a pathogen infection in a cell by administering to  
PT the cell chimeric molecules having a pathogen-detection domain or  
PT pathogen-induced product-detection domain and at least one effector  
PT domain.

XX Example 7; SEQ ID NO 368; 348pp; English.

XX The present invention describes a method for treating or preventing a  
CC pathogen infection in a cell. The method comprises administering to the  
CC cell chimeric molecules (I) having at least one pathogen-detection domain  
CC or pathogen-induced product-detection domain and at least one effector  
CC domain. The pathogen-detection domain or pathogen-induced product-  
CC detection domain is not naturally bound to the effector domain. In the  
CC presence of a pathogen or pathogen-induced product in the cell, the

CC chimeric molecules bind to the pathogen and activate the effector domain.  
 CC Also described: (1) treating or preventing the spread of a pathogen  
 CC infection in an organism; (2) a chimeric molecule (1) having at least one  
 CC pathogen-detection domain or pathogen-induced product-detection domain  
 CC and at least one effector domain; (3) an agent having at least one  
 CC pathogen-interacting or pathogen-induced product-interacting molecular  
 CC structure and at least one effector-mediated product-detection domain  
 CC (1) has antimicrobial activity, and can be used in gene therapy. The  
 CC method is useful for treating or preventing a pathogen infection in a  
 CC cell or organism. The present sequence is used in the exemplification of  
 CC the present invention.  
 XX  
 SQ Sequence 10 AA;  
 Query Match 100.0%; Score 27; DB 7; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GRRGR 5  
 Db |||||  
 1 GRRGR 5  
 RESULT 36  
 ADG44423  
 ID ADG44423 standard; peptide; 10 AA.  
 XX  
 AC ADG44423;  
 XX  
 DT 26-FEB-2004 (first entry)  
 XX  
 DE Anti-pathogen related amino acid sequence SEQ ID NO:222.  
 XX  
 KW pathogen infection; chimeric molecule; pathogen-detection domain;  
 KW pathogen-induced product-detection domain; effector domain;  
 KW antimicrobial; gene therapy.  
 XX  
 OS Synthetic.  
 XX  
 PN WO2003066003-A2.  
 XX  
 PD 14-AUG-2003.  
 XX  
 PF 07-FEB-2003; 2003WO-US003978.  
 XX  
 PR 07-FEB-2002; 2002US-0355022P.  
 XX  
 PR 07-FEB-2002; 2002US-0355359P.  
 XX  
 PR 10-DEC-2002; 2002US-0432386P.  
 XX  
 PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
 XX  
 PI Rider TH;  
 XX  
 DR WPI; 2003-731477/69.  
 XX  
 PT Treating or preventing a pathogen infection in a cell by administering to  
 PT the cell chimeric molecules having a pathogen-detection domain or  
 PT pathogen-induced product-detection domain and at least one effector  
 PT domain.  
 XX  
 PS Example 7; SEQ ID NO 222; 348pp; English.  
 XX  
 CC The present invention describes a method for treating or preventing a  
 CC pathogen infection in a cell. The method comprises administering to the  
 CC cell chimeric molecules (1) having at least one pathogen-detection domain  
 CC or pathogen-induced product-detection domain and at least one effector  
 CC domain. The pathogen-detection domain or pathogen-induced product-  
 CC detection domain is not naturally bound to the effector domain. In the  
 CC presence of a pathogen or pathogen-induced product in the cell, the  
 CC chimeric molecules bind to the pathogen and activate the effector domain.  
 CC Also described: (1) treating or preventing the spread of a pathogen  
 CC infection in an organism; (2) a chimeric molecule (1) having at least one  
 CC pathogen-detection domain or pathogen-induced product-detection domain  
 CC and at least one effector domain; (3) a chimeric molecule (1) having at least one  
 CC pathogen-interacting or pathogen-induced product-interacting molecular

CC pathogen-detection domain or pathogen-induced product-detection domain  
 CC and at least one effector domain; (3) an agent having at least one  
 CC pathogen-interacting or pathogen-induced product-interacting molecular  
 CC structure and at least one effector-mediated product-detection domain  
 CC (4) an assay for detecting a pathogen infection in a cell or organism.  
 CC (1) has antimicrobial activity, and can be used in gene therapy. The  
 CC method is useful for treating or preventing a pathogen infection in a  
 CC cell or organism. The present sequence is used in the exemplification of  
 CC the present invention.  
 XX  
 SQ Sequence 10 AA;  
 Query Match 100.0%; Score 27; DB 7; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GRRGR 5  
 Db |||||  
 1 GRRGR 5  
 RESULT 37  
 ADG44555  
 ID ADG44555 standard; peptide; 10 AA.  
 XX  
 AC ADG44555;  
 XX  
 DT 26-FEB-2004 (first entry)  
 XX  
 DE Anti-pathogen related amino acid sequence SEQ ID NO:354.  
 XX  
 KW pathogen infection; chimeric molecule; pathogen-detection domain;  
 KW pathogen-induced product-detection domain; effector domain;  
 KW antimicrobial; gene therapy.  
 XX  
 OS Synthetic.  
 XX  
 PN WO2003066003-A2.  
 XX  
 PD 14-AUG-2003.  
 XX  
 PF 07-FEB-2003; 2003WO-US003978.  
 XX  
 PR 07-FEB-2002; 2002US-0355022P.  
 XX  
 PR 07-FEB-2002; 2002US-0355359P.  
 XX  
 PR 10-DEC-2002; 2002US-0432386P.  
 XX  
 PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
 XX  
 PI Rider TH;  
 XX  
 DR WPI; 2003-731477/69.  
 XX  
 PT Treating or preventing a pathogen infection in a cell by administering to  
 PT the cell chimeric molecules having a pathogen-detection domain or  
 PT pathogen-induced product-detection domain and at least one effector  
 PT domain.  
 XX  
 PS Example 7; SEQ ID NO 354; 348pp; English.  
 XX  
 CC The present invention describes a method for treating or preventing a  
 CC pathogen infection in a cell. The method comprises administering to the  
 CC cell chimeric molecules (1) having at least one pathogen-detection domain  
 CC or pathogen-induced product-detection domain and at least one effector  
 CC domain. The pathogen-detection domain or pathogen-induced product-  
 CC detection domain is not naturally bound to the effector domain. In the  
 CC presence of a pathogen or pathogen-induced product in the cell, the  
 CC chimeric molecules bind to the pathogen and activate the effector domain.  
 CC Also described: (1) treating or preventing the spread of a pathogen  
 CC infection in an organism; (2) a chimeric molecule (1) having at least one  
 CC pathogen-detection domain or pathogen-induced product-detection domain  
 CC and at least one effector domain; (3) an agent having at least one  
 CC pathogen-interacting or pathogen-induced product-interacting molecular

CC structure and at least one effector-mediating molecular structure; and  
CC (4) an assay for detecting a pathogen infection in a cell or organism.  
CC (1) has antimicrobial activity, and can be used in gene therapy. The  
CC method is useful for treating or preventing a pathogen infection in a  
CC cell or organism. The present sequence is used in the exemplification of  
CC the present invention.  
XX  
SQ Sequence 10 AA;  
  
Query Match 100.0%; Score 27; DB 7; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GRRGR 5  
Db |||||  
1 GRRGR 5  
  
RESULT 38  
ADG44625  
ID ADG44625 standard; peptide; 10 AA.  
XX AC ADG44625;  
XX  
XX 26-FEB-2004 (first entry)  
XX  
DE Anti-pathogen related amino acid sequence SEQ ID NO:424.  
XX  
KW pathogen infection; chimeric molecule; pathogen-detection domain;  
KW pathogen-induced product-detection domain; effector domain;  
KW antimicrobial; gene therapy.  
XX OS Synthetic.  
XX PN WO2003066003-A2.  
XX PD 14-AUG-2003.  
XX PF 07-FEB-2003; 2003WO-US003978.  
XX PR 07-FEB-2002; 2002US-0355022P.  
XX PR 07-FEB-2002; 2002US-0355359P.  
XX PR 10-DEC-2002; 2002US-0432386P.  
XX PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
XX PI Rider TH;  
XX DR WPI; 2003-731477/69.  
XX  
PT Treating or preventing a pathogen infection in a cell by administering to  
PT the cell chimeric molecules having a pathogen-detection domain or  
PT pathogen-induced product-detection domain and at least one effector  
PT domain.  
XX  
PS Example 7; SEQ ID NO 424; 348pp; English.  
XX  
CC The present invention describes a method for treating or preventing a  
CC pathogen infection in a cell. The method comprises administering to the  
CC cell chimeric molecules (1) having at least one pathogen-detection domain  
CC or pathogen-induced product-detection domain and at least one effector  
CC domain. The pathogen-detection domain or pathogen-induced product-  
CC detection domain is not naturally bound to the effector domain. In the  
CC presence of a pathogen or pathogen-induced product in the cell, the  
CC chimeric molecules bind to the pathogen and activate the effector domain.  
CC Also described: (1) treating or preventing the spread of a pathogen  
CC infection in an organism; (2) a chimeric molecule (1) having at least one  
CC pathogen-detection domain or pathogen-induced product-detection domain  
CC and at least one effector domain; (3) an agent having at least one  
CC pathogen-interacting or pathogen-induced product-interacting molecular  
CC structure and at least one effector-mediating molecular structure; and  
CC (4) an assay for detecting a pathogen infection in a cell or organism.  
CC (1) has antimicrobial activity, and can be used in gene therapy. The

CC method is useful for treating or preventing a pathogen infection in a  
CC cell or organism. The present sequence is used in the exemplification of  
CC the present invention.  
XX  
SQ Sequence 10 AA;  
  
Query Match 100.0%; Score 27; DB 7; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GRRGR 5  
Db |||||  
1 GRRGR 5  
  
RESULT 39  
ADG44541  
ID ADG44541 standard; peptide; 10 AA.  
XX AC ADG44541;  
XX  
XX 26-FEB-2004 (first entry)  
XX  
DE Anti-pathogen related amino acid sequence SEQ ID NO:340.  
XX  
KW pathogen infection; chimeric molecule; pathogen-detection domain;  
KW pathogen-induced product-detection domain; effector domain;  
KW antimicrobial; gene therapy.  
XX OS Synthetic.  
XX PN WO2003066003-A2.  
XX PD 14-AUG-2003.  
XX PF 07-FEB-2003; 2003WO-US003978.  
XX PR 07-FEB-2002; 2002US-0355022P.  
XX PR 07-FEB-2002; 2002US-0355359P.  
XX PR 10-DEC-2002; 2002US-0432386P.  
XX PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
XX PI Rider TH;  
XX DR WPI; 2003-731477/69.  
XX  
PT Treating or preventing a pathogen infection in a cell by administering to  
PT the cell chimeric molecules having a pathogen-detection domain or  
PT pathogen-induced product-detection domain and at least one effector  
PT domain.  
XX  
PS Example 7; SEQ ID NO 340; 348pp; English.  
XX  
CC The present invention describes a method for treating or preventing a  
CC pathogen infection in a cell. The method comprises administering to the  
CC cell chimeric molecules (1) having at least one pathogen-detection domain  
CC or pathogen-induced product-detection domain and at least one effector  
CC domain. The pathogen-detection domain or pathogen-induced product-  
CC detection domain is not naturally bound to the effector domain. In the  
CC presence of a pathogen or pathogen-induced product in the cell, the  
CC chimeric molecules bind to the pathogen and activate the effector domain.  
CC Also described: (1) treating or preventing the spread of a pathogen  
CC infection in an organism; (2) a chimeric molecule (1) having at least one  
CC pathogen-detection domain or pathogen-induced product-detection domain  
CC and at least one effector domain; (3) an agent having at least one  
CC pathogen-interacting or pathogen-induced product-interacting molecular  
CC structure and at least one effector-mediating molecular structure; and  
CC (4) an assay for detecting a pathogen infection in a cell or organism.  
CC (1) has antimicrobial activity, and can be used in gene therapy. The  
CC method is useful for treating or preventing a pathogen infection in a  
CC cell or organism. The present sequence is used in the exemplification of  
CC the present invention.

XX SQ Sequence 10 AA;  
 Query Match 100.0%; Score 27; DB 7; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02; Indels 0; Gaps 0;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
 |||||  
 Db 1 GRGR 5

RESULT 40  
 ADG4471  
 ID ADG4471 standard; peptide; 10 AA.  
 XX  
 AC ADG4471;  
 XX  
 DT 26-FEB-2004 (first entry)  
 XX  
 DE Anti-pathogen related amino acid sequence SEQ ID NO:270.  
 XX  
 KW pathogen infection; chimeric molecule; pathogen-detection domain;  
 KW pathogen-induced product-detection domain; effector domain;  
 KW antimicrobial; gene therapy.  
 XX  
 OS Synthetic.  
 XX  
 PN WO2003066003-A2.  
 XX  
 PD 14-AUG-2003.  
 XX  
 PF 07-FEB-2003; 2003WO-US003978.  
 XX  
 PR 07-FEB-2002; 2002US-0355022P.  
 PR 07-FEB-2002; 2002US-0355359P.  
 PR 10-DEC-2002; 2002US-0432386P.  
 XX  
 PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
 XX  
 PI Rider TH;  
 XX  
 DR WPI; 2003-731477/69.  
 XX  
 PT Treating or preventing a pathogen infection in a cell by administering to  
 PT the cell chimeric molecules having a pathogen-detection domain or  
 PT pathogen-induced product-detection domain and at least one effector  
 PT domain.  
 XX  
 PS Example 7; SEQ ID NO 270; 348pp; English.  
 XX  
 CC The present invention describes a method for treating or preventing a  
 CC pathogen infection in a cell. The method comprises administering to the  
 CC cell chimeric molecules (1) having at least one pathogen-detection domain  
 CC or pathogen-induced product-detection domain and at least one effector  
 CC domain. The pathogen-detection domain or pathogen-induced product-  
 CC detection domain is not naturally bound to the effector domain. In the  
 CC presence of a pathogen or pathogen-induced product in the cell, the  
 CC chimeric molecules bind to the pathogen and activate the effector domain.  
 CC Also described: (1) treating or preventing the spread of a pathogen  
 CC infection in an organism; (2) a chimeric molecule (1) having at least one  
 CC pathogen-detection domain or pathogen-induced product-detection domain  
 CC and at least one effector domain; (3) an agent having at least one  
 CC pathogen-interacting or pathogen-induced product-interacting molecular  
 CC structure and at least one effector-mediated molecular structure; and  
 CC (4) an assay for detecting a pathogen infection in a cell or organism.  
 CC (1) has antimicrobial activity, and can be used in gene therapy. The  
 CC method is useful for treating or preventing a pathogen infection in a  
 CC cell or organism. The present sequence is used in the exemplification of  
 CC the present invention.  
 XX  
 SQ Sequence 10 AA;

Query Match 100.0%; Score 27; DB 7; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02; Indels 0; Gaps 0;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
 |||||  
 Db 1 GRGR 5

RESULT 41  
 ADG4583  
 ID ADG4583 standard; peptide; 10 AA.  
 XX  
 AC ADG4583;  
 XX  
 DT 26-FEB-2004 (first entry)  
 XX  
 DE Anti-pathogen related amino acid sequence SEQ ID NO:382.  
 XX  
 KW pathogen infection; chimeric molecule; pathogen-detection domain;  
 KW pathogen-induced product-detection domain; effector domain;  
 KW antimicrobial; gene therapy.  
 XX  
 OS Synthetic.  
 XX  
 PN WO2003066003-A2.  
 XX  
 PD 14-AUG-2003.  
 XX  
 PF 07-FEB-2003; 2003WO-US003978.  
 XX  
 PR 07-FEB-2002; 2002US-0355022P.  
 PR 07-FEB-2002; 2002US-0355359P.  
 PR 10-DEC-2002; 2002US-0432386P.  
 XX  
 PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
 XX  
 PI Rider TH;  
 XX  
 DR WPI; 2003-731477/69.  
 XX  
 PT Treating or preventing a pathogen infection in a cell by administering to  
 PT the cell chimeric molecules having a pathogen-detection domain or  
 PT pathogen-induced product-detection domain and at least one effector  
 PT domain.  
 XX  
 PS Example 7; SEQ ID NO 382; 348pp; English.  
 XX  
 CC The present invention describes a method for treating or preventing a  
 CC pathogen infection in a cell. The method comprises administering to the  
 CC cell chimeric molecules (1) having at least one pathogen-detection domain  
 CC or pathogen-induced product-detection domain and at least one effector  
 CC domain. The pathogen-detection domain or pathogen-induced product-  
 CC detection domain is not naturally bound to the effector domain. In the  
 CC presence of a pathogen or pathogen-induced product in the cell, the  
 CC chimeric molecules bind to the pathogen and activate the effector domain.  
 CC Also described: (1) treating or preventing the spread of a pathogen  
 CC infection in an organism; (2) a chimeric molecule (1) having at least one  
 CC pathogen-detection domain or pathogen-induced product-detection domain  
 CC and at least one effector domain; (3) an agent having at least one  
 CC pathogen-interacting or pathogen-induced product-interacting molecular  
 CC structure and at least one effector-mediated molecular structure; and  
 CC (4) an assay for detecting a pathogen infection in a cell or organism.  
 CC (1) has antimicrobial activity, and can be used in gene therapy. The  
 CC method is useful for treating or preventing a pathogen infection in a  
 CC cell or organism. The present sequence is used in the exemplification of  
 CC the present invention.  
 XX  
 SQ Sequence 10 AA;

Query Match 100.0%; Score 27; DB 7; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02; Indels 0; Gaps 0;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
Db 1 GRRGR 5

RESULT 42  
ADG44457  
ID ADG44457 standard; peptide; 10 AA.  
XX  
AC ADG44457;  
XX  
DT 26-FEB-2004 (first entry)  
XX  
DE Anti-pathogen related amino acid sequence SEQ ID NO:256.  
XX  
KW pathogen infection; chimeric molecule; pathogen-detection domain;  
KW pathogen-induced product-detection domain; effector domain;  
KW antimicrobial; gene therapy.  
XX  
OS Synthetic.  
XX  
PN WO2003066003-A2.  
XX  
PD 14-AUG-2003.  
XX  
PF 07-FEB-2003; 2003WO-US003978.  
XX  
PR 07-FEB-2002; 2002US-0355022P.  
PR 07-FEB-2002; 2002US-0355359P.  
PR 10-DEC-2002; 2002US-0432386P.  
XX  
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
XX  
PI Rider TH;  
XX  
DR WPI; 2003-731477/69.  
XX  
PT Treating or preventing a pathogen infection in a cell by administering to  
PT the cell chimeric molecules having a pathogen-detection domain or  
PT pathogen-induced product-detection domain and at least one effector  
PT domain.  
XX  
PS Example 7; SEQ ID NO 256; 348pp; English.  
XX  
CC The present invention describes a method for treating or preventing a  
CC pathogen infection in a cell. The method comprises administering to the  
CC cell chimeric molecules (I) having at least one pathogen-detection domain  
CC or pathogen-induced product-detection domain and at least one effector  
CC domain. The pathogen-detection domain or pathogen-induced product-  
CC detection domain is not naturally bound to the effector domain. In the  
CC presence of a pathogen or pathogen-induced product in the cell, the  
CC chimeric molecules bind to the pathogen and activate the effector domain.  
CC Also described: (1) treating or preventing the spread of a pathogen  
CC infection in an organism; (2) a chimeric molecule (I) having at least one  
CC pathogen-detection domain or pathogen-induced product-detection domain  
CC and at least one effector domain; (3) an agent having at least one  
CC pathogen-interacting or pathogen-mediated product-interacting molecular  
CC structure and at least one effector-mediated product-interacting molecular  
CC structure and at least one effector-mediated product-interacting molecular  
CC (4) an assay for detecting a pathogen infection in a cell or organism.  
CC (I) has antimicrobial activity, and can be used in gene therapy. The  
CC method is useful for treating or preventing a pathogen infection in a  
CC cell or organism. The present sequence is used in the exemplification of  
CC the present invention.  
XX  
SQ Sequence 10 AA;

Query Match 100.0%; Score 27; DB 7; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
Db 1 GRRGR 5

Db 1 GRRGR 5

RESULT 43  
ADG44667  
ID ADG44667 standard; peptide; 10 AA.  
XX  
AC ADG44667;  
XX  
DT 26-FEB-2004 (first entry)  
XX  
DE Anti-pathogen related amino acid sequence SEQ ID NO:466.  
XX  
KW pathogen infection; chimeric molecule; pathogen-detection domain;  
KW pathogen-induced product-detection domain; effector domain;  
KW antimicrobial; gene therapy.  
XX  
OS Synthetic.  
XX  
PN WO2003066003-A2.  
XX  
PD 14-AUG-2003.  
XX  
PF 07-FEB-2003; 2003WO-US003978.  
XX  
PR 07-FEB-2002; 2002US-0355022P.  
PR 07-FEB-2002; 2002US-0355359P.  
PR 10-DEC-2002; 2002US-0432386P.  
XX  
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
XX  
PI Rider TH;  
XX  
DR WPI; 2003-731477/69.  
XX  
PT Treating or preventing a pathogen infection in a cell by administering to  
PT the cell chimeric molecules having a pathogen-detection domain or  
PT pathogen-induced product-detection domain and at least one effector  
PT domain.  
XX  
PS Example 7; SEQ ID NO 466; 348pp; English.  
XX  
CC The present invention describes a method for treating or preventing a  
CC pathogen infection in a cell. The method comprises administering to the  
CC cell chimeric molecules (I) having at least one pathogen-detection domain  
CC or pathogen-induced product-detection domain and at least one effector  
CC domain. The pathogen-detection domain or pathogen-induced product-  
CC detection domain is not naturally bound to the effector domain. In the  
CC presence of a pathogen or pathogen-induced product in the cell, the  
CC chimeric molecules bind to the pathogen and activate the effector domain.  
CC Also described: (1) treating or preventing the spread of a pathogen  
CC infection in an organism; (2) a chimeric molecule (I) having at least one  
CC pathogen-detection domain or pathogen-induced product-detection domain  
CC and at least one effector domain; (3) an agent having at least one  
CC pathogen-interacting or pathogen-mediated product-interacting molecular  
CC structure and at least one effector-mediated product-interacting molecular  
CC (4) an assay for detecting a pathogen infection in a cell or organism.  
CC (I) has antimicrobial activity, and can be used in gene therapy. The  
CC method is useful for treating or preventing a pathogen infection in a  
CC cell or organism. The present sequence is used in the exemplification of  
CC the present invention.  
XX  
SQ Sequence 10 AA;

Query Match 100.0%; Score 27; DB 7; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
Db 1 GRRGR 5



```

RESULT 44
ADG44440
ID ADG44440 standard; peptide; 10 AA.
XX
XX
AC ADG44440;
XX
XX
DT 26-FEB-2004 (first entry)
XX
XX
DE Anti-pathogen related amino acid sequence SEQ ID NO:239.
XX
XX
DE pathogen infection; chimeric molecule; pathogen-detection domain;
XX
KW pathogen-induced product-detection domain; effector domain;
KW
KW antimicrobial; gene therapy.
XX
OS Synthetic.
XX
XX WO2003066003-A2.
XX
XX PD 14-AUG-2003.
XX
XX PF 07-FEB-2003; 2003WO-US003978.
XX
XX PF 07-FEB-2002; 2002US-0355022P.
XX
XX PR 07-FEB-2002; 2002US-0355359P.
XX
XX PR 10-DEC-2002; 2002US-0432386P.
XX
XX PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX
XX PI Rider TH;
XX
XX DR WPI; 2003-731477/69.
XX
XX PT Treating or preventing a pathogen infection in a cell by administering to
XX
PT the cell chimeric molecules having a pathogen-detection domain or
PT
PT pathogen-induced product-detection domain and at least one effector
PT
PT domain.
XX
XX Example 7; SEQ ID NO 239; 348pp; English.
XX
XX The present invention describes a method for treating or preventing a
XX
CC pathogen infection in a cell. The method comprises administering to the
CC
CC cell chimeric molecules (1) having at least one pathogen-detection domain
CC
CC or pathogen-induced product-detection domain and at least one effector
CC
CC domain. The pathogen-detection domain or pathogen-induced product-
CC
CC detection domain is not naturally bound to the effector domain. In the
CC
CC presence of a pathogen or pathogen-induced product in the cell, the
CC
CC chimeric molecules bind to the pathogen and activate the effector domain.
CC
CC Also described: (1) treating or preventing the spread of a pathogen
CC
CC infection in an organism; (2) a chimeric molecule (1) having at least one
CC
CC pathogen-detection domain or pathogen-induced product-detection domain
CC
CC and at least one effector domain; (3) an agent having at least one
CC
CC pathogen-detection domain or pathogen-induced product-detection domain
CC
CC structure and at least one effector-mediated product-interacting molecular
CC
CC and at least one effector domain; (3) an agent having at least one
CC
CC pathogen-interacting or pathogen-induced product-interacting molecular
CC
CC structure and at least one effector-mediated product-interacting molecular
CC
CC (4) an assay for detecting a pathogen infection in a cell or organism.
CC
CC (1) has antimicrobial activity, and can be used in gene therapy. The
CC
CC method is useful for treating or preventing a pathogen infection in a
CC
CC cell or organism. The present sequence is used in the exemplification of
CC
CC the present invention.
XX
XX Sequence 10 AA;
XX
SQ Query Match 100.0%; Score 27; DB 7; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GRRGR 5
DB |||||
1 GRRGR 5
RESULT 45
ADG44485
ID ADG44485 standard; peptide; 10 AA.
XX
XX
AC ADG44485;
XX
XX
DT 26-FEB-2004 (first entry)
XX
XX
DE Anti-pathogen related amino acid sequence SEQ ID NO:284.
XX
XX
DE pathogen infection; chimeric molecule; pathogen-detection domain;
XX
KW pathogen-induced product-detection domain; effector domain;
KW
KW antimicrobial; gene therapy.
XX
OS Synthetic.
XX
XX WO2003066003-A2.
XX
XX PD 14-AUG-2003.
XX
XX PF 07-FEB-2003; 2003WO-US003978.
XX
XX PF 07-FEB-2002; 2002US-0355022P.
XX
XX PR 07-FEB-2002; 2002US-0355359P.
XX
XX PR 10-DEC-2002; 2002US-0432386P.
XX
XX PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX
XX PI Rider TH;
XX
XX DR WPI; 2003-731477/69.
XX
XX PT Treating or preventing a pathogen infection in a cell by administering to
XX
PT the cell chimeric molecules having a pathogen-detection domain or
PT
PT pathogen-induced product-detection domain and at least one effector
PT
PT domain.
XX
XX Example 7; SEQ ID NO 284; 348pp; English.
XX
XX The present invention describes a method for treating or preventing a
XX
CC pathogen infection in a cell. The method comprises administering to the
CC
CC cell chimeric molecules (1) having at least one pathogen-detection domain
CC
CC or pathogen-induced product-detection domain and at least one effector
CC
CC domain. The pathogen-detection domain or pathogen-induced product-
CC
CC detection domain is not naturally bound to the effector domain. In the
CC
CC presence of a pathogen or pathogen-induced product in the cell, the
CC
CC chimeric molecules bind to the pathogen and activate the effector domain.
CC
CC Also described: (1) treating or preventing the spread of a pathogen
CC
CC infection in an organism; (2) a chimeric molecule (1) having at least one
CC
CC pathogen-detection domain or pathogen-induced product-detection domain
CC
CC structure and at least one effector-mediated product-interacting molecular
CC
CC and at least one effector domain; (3) an agent having at least one
CC
CC pathogen-interacting or pathogen-induced product-interacting molecular
CC
CC structure and at least one effector-mediated product-interacting molecular
CC
CC (4) an assay for detecting a pathogen infection in a cell or organism.
CC
CC (1) has antimicrobial activity, and can be used in gene therapy. The
CC
CC method is useful for treating or preventing a pathogen infection in a
CC
CC cell or organism. The present sequence is used in the exemplification of
CC
CC the present invention.
XX
XX Sequence 10 AA;
XX
SQ Query Match 100.0%; Score 27; DB 7; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GRRGR 5
DB |||||
1 GRRGR 5
RESULT 46
ADG44513
ID ADG44513 standard; peptide; 10 AA.
XX
XX
AC ADG44513;
XX
XX
DT 26-FEB-2004 (first entry)
XX
XX
DE Anti-pathogen related amino acid sequence SEQ ID NO:284.
XX
XX
DE pathogen infection; chimeric molecule; pathogen-detection domain;
XX
KW pathogen-induced product-detection domain; effector domain;
KW
KW antimicrobial; gene therapy.
XX
OS Synthetic.
XX
XX WO2003066003-A2.
XX
XX PD 14-AUG-2003.
XX
XX PF 07-FEB-2003; 2003WO-US003978.
XX
XX PF 07-FEB-2002; 2002US-0355022P.
XX
XX PR 07-FEB-2002; 2002US-0355359P.
XX
XX PR 10-DEC-2002; 2002US-0432386P.
XX
XX PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX
XX PI Rider TH;
XX
XX DR WPI; 2003-731477/69.
XX
XX PT Treating or preventing a pathogen infection in a cell by administering to
XX
PT the cell chimeric molecules having a pathogen-detection domain or
PT
PT pathogen-induced product-detection domain and at least one effector
PT
PT domain.
XX
XX Example 7; SEQ ID NO 284; 348pp; English.
XX
XX The present invention describes a method for treating or preventing a
XX
CC pathogen infection in a cell. The method comprises administering to the
CC
CC cell chimeric molecules (1) having at least one pathogen-detection domain
CC
CC or pathogen-induced product-detection domain and at least one effector
CC
CC domain. The pathogen-detection domain or pathogen-induced product-
CC
CC detection domain is not naturally bound to the effector domain. In the
CC
CC presence of a pathogen or pathogen-induced product in the cell, the
CC
CC chimeric molecules bind to the pathogen and activate the effector domain.
CC
CC Also described: (1) treating or preventing the spread of a pathogen
CC
CC infection in an organism; (2) a chimeric molecule (1) having at least one
CC
CC pathogen-detection domain or pathogen-induced product-detection domain
CC
CC structure and at least one effector-mediated product-interacting molecular
CC
CC and at least one effector domain; (3) an agent having at least one
CC
CC pathogen-interacting or pathogen-induced product-interacting molecular
CC
CC structure and at least one effector-mediated product-interacting molecular
CC
CC (4) an assay for detecting a pathogen infection in a cell or organism.
CC
CC (1) has antimicrobial activity, and can be used in gene therapy. The
CC
CC method is useful for treating or preventing a pathogen infection in a
CC
CC cell or organism. The present sequence is used in the exemplification of
CC
CC the present invention.
XX
XX Sequence 10 AA;
XX
SQ Query Match 100.0%; Score 27; DB 7; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GRRGR 5
DB |||||
1 GRRGR 5

```

DT 26-FEB-2004 (first entry)  
XX Anti-pathogen related amino acid sequence SEQ ID NO:312.  
DE pathogen infection; chimeric molecule; pathogen-detection domain;  
XX pathogen-induced product-detection domain; effector domain;  
KW antimicrobial; gene therapy.  
KW Synthetic.  
XX WO2003066003-A2.  
OS 14-AUG-2003.  
XX 07-FEB-2003; 2003WO-US003978.  
PN 07-FEB-2002; 2002US-0355022P.  
XX 07-FEB-2002; 2002US-0355359P.  
PF 10-DEC-2002; 2002US-0432386P.  
XX (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
PA Rider TH;  
XX WPI; 2003-731477/69.  
PI Treating or preventing a pathogen infection in a cell by administering to  
XX the cell chimeric molecules having a pathogen-detection domain or  
XX pathogen-induced product-detection domain and at least one effector  
XX domain.  
PS Example 7; SEQ ID NO 312; 348pp; English.  
XX The present invention describes a method for treating or preventing a  
CC pathogen infection in a cell. The method comprises administering to the  
CC cell chimeric molecules (I) having at least one pathogen-detection domain  
CC or pathogen-induced product-detection domain and at least one effector  
CC domain. The pathogen-detection domain or pathogen-induced product-  
CC detection domain is not naturally bound to the effector domain. In the  
CC presence of a pathogen or pathogen-induced product in the cell, the  
CC chimeric molecules bind to the pathogen and activate the effector domain.  
CC Also described: (1) treating or preventing the spread of a pathogen  
CC infection in an organism; (2) a chimeric molecule (I) having at least one  
CC pathogen-detection domain or pathogen-induced product-detection domain  
CC and at least one effector domain; (3) an agent having at least one  
CC pathogen-interacting or pathogen-induced product-interacting molecular  
CC structure and at least one effector-mediated product-interacting molecular  
CC structure and at least one effector-mediated product-interacting molecular  
CC (4) an assay for detecting a pathogen infection in a cell or organism.  
CC (I) has antimicrobial activity, and can be used in gene therapy. The  
CC method is useful for treating or preventing a pathogen infection in a  
CC cell or organism. The present sequence is used in the exemplification of  
CC the present invention.  
XX Sequence 10 AA;  
SQ Query Match 100.0%; Score 27; DB 7; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GRRGR 5  
Db 1 GRRGR 5  
RESULT 47  
ADG44527  
ID ADG44527 standard; peptide; 10 AA.  
XX AC ADG44527;  
AC 26-FEB-2004 (first entry)  
XX 26-FEB-2004 (first entry)  
DT Anti-pathogen related amino acid sequence SEQ ID NO:326.  
XX Anti-pathogen related amino acid sequence SEQ ID NO:326.  
DE

XX pathogen infection; chimeric molecule; pathogen-detection domain;  
KW pathogen-induced product-detection domain; effector domain;  
KW antimicrobial; gene therapy.  
XX Synthetic.  
XX WO2003066003-A2.  
PN 14-AUG-2003.  
XX 07-FEB-2003; 2003WO-US003978.  
PF 07-FEB-2002; 2002US-0355022P.  
XX 07-FEB-2002; 2002US-0355359P.  
PR 10-DEC-2002; 2002US-0432386P.  
XX (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
PA Rider TH;  
XX WPI; 2003-731477/69.  
PI Treating or preventing a pathogen infection in a cell by administering to  
XX the cell chimeric molecules having a pathogen-detection domain or  
XX pathogen-induced product-detection domain and at least one effector  
XX domain.  
PS Example 7; SEQ ID NO 326; 348pp; English.  
XX The present invention describes a method for treating or preventing a  
CC pathogen infection in a cell. The method comprises administering to the  
CC cell chimeric molecules (I) having at least one pathogen-detection domain  
CC or pathogen-induced product-detection domain and at least one effector  
CC domain. The pathogen-detection domain or pathogen-induced product-  
CC detection domain is not naturally bound to the effector domain. In the  
CC presence of a pathogen or pathogen-induced product in the cell, the  
CC chimeric molecules bind to the pathogen and activate the effector domain.  
CC Also described: (1) treating or preventing the spread of a pathogen  
CC infection in an organism; (2) a chimeric molecule (I) having at least one  
CC pathogen-detection domain or pathogen-induced product-detection domain  
CC and at least one effector domain; (3) an agent having at least one  
CC pathogen-interacting or pathogen-induced product-interacting molecular  
CC structure and at least one effector-mediated product-interacting molecular  
CC structure and at least one effector-mediated product-interacting molecular  
CC (4) an assay for detecting a pathogen infection in a cell or organism.  
CC (I) has antimicrobial activity, and can be used in gene therapy. The  
CC method is useful for treating or preventing a pathogen infection in a  
CC cell or organism. The present sequence is used in the exemplification of  
CC the present invention.  
XX Sequence 10 AA;  
SQ Query Match 100.0%; Score 27; DB 7; Length 10;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GRRGR 5  
Db 1 GRRGR 5  
RESULT 48  
ADG44639  
ID ADG44639 standard; peptide; 10 AA.  
XX AC ADG44639;  
AC 26-FEB-2004 (first entry)  
XX 26-FEB-2004 (first entry)  
DT Anti-pathogen related amino acid sequence SEQ ID NO:438.  
XX Anti-pathogen related amino acid sequence SEQ ID NO:438.  
DE pathogen infection; chimeric molecule; pathogen-detection domain;  
KW pathogen-induced product-detection domain; effector domain;  
KW

```

KW antimicrobial; gene therapy.
XX
OS Synthetic.
XX
PN WO2003066003-A2.
XX
PD 14-AUG-2003.
XX
PF 07-FEB-2003; 2003WO-US003978.
XX
PR 07-FEB-2002; 2002US-0355022P.
XX
PR 07-FEB-2002; 2002US-0355359P.
XX
PR 10-DEC-2002; 2002US-0432386P.
XX
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX
PI Rider TH;
XX
PI WPI; 2003-731477/69.
XX
DR Treating or preventing a pathogen infection in a cell by administering to
XX the cell chimeric molecules having a pathogen-detection domain or
XX pathogen-induced product-detection domain and at least one effector
XX domain.
XX
XX Example 7; SEQ ID NO 438; 348pp; English.
XX
CC The present invention describes a method for treating or preventing a
CC pathogen infection in a cell. The method comprises administering to the
CC cell chimeric molecules (1) having at least one pathogen-detection domain
CC or pathogen-induced product-detection domain and at least one effector
CC domain. The pathogen-detection domain or pathogen-induced product-
CC detection domain is not naturally bound to the effector domain. In the
CC presence of a pathogen or pathogen-induced product in the cell, the
CC chimeric molecules bind to the pathogen and activate the spread of a pathogen.
CC Also described: (1) treating or preventing the spread of a pathogen
CC infection in an organism; (2) a chimeric molecule (1) having at least one
CC pathogen-detection domain or pathogen-induced product-detection domain
CC and at least one effector domain; (3) an agent having at least one
CC pathogen-detection domain or pathogen-induced product-detection domain
CC structure and at least one effector-mediated product-interacting molecular
CC (4) an assay for detecting a pathogen infection in a cell or organism.
CC (1) has antimicrobial activity, and can be used in gene therapy. The
CC method is useful for treating or preventing a pathogen infection in a
CC cell or organism. The present sequence is used in the exemplification of
CC the present invention.
XX
SQ Sequence 10 AA;
Query Match 100.0%; Score 27; DB 7; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GRRGR 5
Db |||||
1 GRRGR 5
RESULT 49
ADG44653
ID ADG44653 standard; peptide; 10 AA.
XX
AC ADG44653;
XX
DT 26-FEB-2004 (first entry)
XX
DE Anti-pathogen related amino acid sequence SEQ ID NO:452.
XX
KW pathogen infection; chimeric molecule; pathogen-detection domain;
KW pathogen-induced product-detection domain; effector domain;
KW antimicrobial; gene therapy.
XX
OS Synthetic.
XX

```

---

```

XX
PN WO2003066003-A2.
XX
PD 14-AUG-2003.
XX
PF 07-FEB-2003; 2003WO-US003978.
XX
PR 07-FEB-2002; 2002US-0355022P.
XX
PR 07-FEB-2002; 2002US-0355359P.
XX
PR 10-DEC-2002; 2002US-0432386P.
XX
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX
PI Rider TH;
XX
PI WPI; 2003-731477/69.
XX
DR Treating or preventing a pathogen infection in a cell by administering to
XX the cell chimeric molecules having a pathogen-detection domain or
XX pathogen-induced product-detection domain and at least one effector
XX domain.
XX
XX Example 7; SEQ ID NO 452; 348pp; English.
XX
CC The present invention describes a method for treating or preventing a
CC pathogen infection in a cell. The method comprises administering to the
CC cell chimeric molecules (1) having at least one pathogen-detection domain
CC or pathogen-induced product-detection domain and at least one effector
CC domain. The pathogen-detection domain or pathogen-induced product-
CC detection domain is not naturally bound to the effector domain. In the
CC presence of a pathogen or pathogen-induced product in the cell, the
CC chimeric molecules bind to the pathogen and activate the spread of a pathogen.
CC Also described: (1) treating or preventing the spread of a pathogen
CC infection in an organism; (2) a chimeric molecule (1) having at least one
CC pathogen-detection domain or pathogen-induced product-detection domain
CC and at least one effector domain; (3) an agent having at least one
CC pathogen-detection domain or pathogen-induced product-interacting molecular
CC structure and at least one effector-mediated product-interacting molecular
CC (4) an assay for detecting a pathogen infection in a cell or organism.
CC (1) has antimicrobial activity, and can be used in gene therapy. The
CC method is useful for treating or preventing a pathogen infection in a
CC cell or organism. The present sequence is used in the exemplification of
CC the present invention.
XX
SQ Sequence 10 AA;
Query Match 100.0%; Score 27; DB 7; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GRRGR 5
Db |||||
1 GRRGR 5
RESULT 50
ADG44597
ID ADG44597 standard; peptide; 10 AA.
XX
AC ADG44597;
XX
DT 26-FEB-2004 (first entry)
XX
DE Anti-pathogen related amino acid sequence SEQ ID NO:396.
XX
KW pathogen infection; chimeric molecule; pathogen-detection domain;
KW pathogen-induced product-detection domain; effector domain;
KW antimicrobial; gene therapy.
XX
OS Synthetic.
XX
PN WO2003066003-A2.
XX

```

XX	07-FEB-2002; 2002US-0355022P.
XX	07-FEB-2002; 2002US-0355359P.
XX	10-DEC-2002; 2002US-0432386P.
XX	(MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX	Rider TH;
XX	WPI; 2003-731477/69.
XX	Example 7; SEQ ID NO 298; 348pp; English.
XX	The present invention describes a method for treating or preventing a
XX	pathogen infection in a cell. The method comprises administering to the
XX	cell chimeric molecules (I) having at least one pathogen-detection domain
XX	or pathogen-induced product-detection domain and at least one effector
XX	domain. The pathogen-detection domain or pathogen-induced product-
XX	detection domain is not naturally bound to the effector domain. In the
XX	presence of a pathogen or pathogen-induced product in the cell, the
XX	chimeric molecules bind to the pathogen and activate the effector domain.
XX	Also described: (1) treating or preventing the spread of a pathogen
XX	infection in an organism; (2) a chimeric molecule (I) having at least one
XX	pathogen-detection domain or pathogen-induced product-detection domain
XX	and at least one effector domain; (3) an agent having at least one
XX	pathogen-interacting or pathogen-induced product-interacting molecular
XX	structure and at least one effector-mediating molecular structure; and
XX	(4) an assay for detecting a pathogen infection in a cell or organism.
XX	(I) has antimicrobial activity, and can be used in gene therapy. The
XX	method is useful for treating or preventing a pathogen infection in a
XX	cell or organism. The present sequence is used in the exemplification of
XX	the present invention.
XX	Sequence 10 AA;
XX	Query Match 100.0%; Score 27; DB 7; Length 10;
XX	Best Local Similarity 100.0%; Pred. No. 1.9e+02;
XX	Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0
Qy	1 GRRGR 5
Db	1 GRRGR 5
RESULT 52	
AAR98451	
ID	AAR98451 standard; peptide; 11 AA.
XX	AAR98451;
XX	12-FEB-1997 (first entry)
XX	Peptide with cell attachment promoting activity.
XX	Hyaluronate; treatment; wound healing; burns; skin graft; ulcer;
KW	tissue regeneration; matrix; matrices; fibronectin; cancer; osteoporosis;
KW	thrombosis.
XX	Synthetic.
XX	Key Location/Qualifiers
FH	11
FT	Modified-site
FT	/note= "Amidated"
XX	W09620002-A1.
XX	04-JUL-1996.
XX	

```

PF 21-DEC-1995; 95WO-US016959.
XX
PR 23-DEC-1994; 94US-00363213.
XX
PA (LJOL-) LA JOLLA CANCER RES FOUND.
XX
PI Dickerson KT, Glass JR, Liu L, Polarek JW, Craig WS, Mullen DG;
PI Cheng S;
XX
DR WPI; 1996-321641/32.
XX
XX Crosslinked hyaluronate-RGD peptide conjugates - prepd. by epoxide,
PT sodium periodate or tresyl chloride methods, provide temporary matrix for
PT wound healing and tissue regeneration.
XX
PS Claim 42; Page 35; 48pp; English.
XX
CC Compositions comprising a cross linked hyaluronate polymer and a peptide
CC having cell attachment promoting activity can be used for treating wounds
CC such as severe burns, skin graft donor sites, decubitus ulcers, diabetic
CC ulcers, surgical incisions and keloid- forming wounds. They can also be
CC used for inducing tissue regeneration. The conjugate acts as a temporary
CC replacement matrix that encourages cell migration into the wound and
CC speeds healing. As the wound heals, the conjugate is slowly broken down
CC by the migrating cells and is replaced by a natural matrix. The peptides
CC contain the amino acid sequence Y-Gly-Asp, where Y is Arg or D-Arg, at
CC least two more amino acids selected from (D-)Arg, (D-)Lys, (D-)Orn and (D
CC -)HomoArg. The peptides can be used to competitively inhibit the binding
CC of cells to RGD containing adhesive proteins such as fibronectin for the
CC treatment of e.g cancer, osteoporosis or thrombosis. Peptides are
CC described in AAR98433-36, AAR98438-43 and AAR98445-53
XX
SQ Sequence 11 AA;

Query Match 100.0%; Score 27; DB 2; Length 11;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5
Db |||||
1 GRRGR 5

RESULT 53
AAW39793
ID AAW39793 standard; protein; 11 AA.
XX
XX AAW39793;
XX
DT 11-JUN-1998 (first entry)
XX
DE Tobacco PABFC repeated AT hook.
XX
KW Palindromic element binding factor; PABF; tobacco; cis-acting element;
KW transcription enhancer; heterologous promoter; AAT repeat element;
KW transcription factor; AT hook.
XX
OS Nicotiana tabacum.
XX
PN WO9749727-A1.
XX
PD 31-DEC-1997.
XX
PF 27-JUN-1997; 97WO-US011156.
XX
PR 27-JUN-1996; 96US-00669721.
XX
PA (SALK ) SALK INST BIOLOGICAL STUDIES.
XX
PI Lamb CJ, Doerner P, Laible G;
XX
DR WPI; 1998-077110/07.
XX
XX New isolated enhancer and transcription factor - used for increasing the
PT recombinant expression of proteins, particularly in plants, e.g. for
PT increasing production or providing pest resistance.
XX
PS Disclosure; Fig 7B; 65pp; English.
XX
AAW39790-W39796 are AT hook motifs used in the characterisation of a
CC novel tobacco palindromic element binding factor, (PABF). This PABF binds
CC to the sequence (AATT)n where n at least 2. The (AATT)n sequence has cis-
CC acting, non-specific enhancer activity. It can be linked to a
CC heterologous promoter operably linked with a gene to increase expression
CC of the gene in a cell, particularly in plants. It can provide for
CC increased expression of proteins such as nutritionally important
CC proteins, growth promoting factors, proteins for early flowering in
CC plants, proteins giving protection to the plant under certain
CC environmental conditions, e.g. proteins conferring resistance to metals
CC or other toxic substances, such as herbicides or pesticides, stress-
CC related proteins which confer tolerance to temperature extremes, proteins
CC conferring resistance to fungi, bacteria, viruses, insects and nematodes,
CC proteins of specific commercial value, e.g. enzymes involved in metabolic
CC pathways, such as EPSP synthase. The PABF polypeptides act as
CC transcription factor and bind to the (AATT) repeat element to further
CC boost the activity of the enhancer
XX
SQ Sequence 11 AA;

Query Match 100.0%; Score 27; DB 2; Length 11;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5
Db |||||
4 GRRGR 8

RESULT 54
AAW39794
ID AAW39794 standard; protein; 11 AA.
XX
XX AAW39794;
XX
DT 11-JUN-1998 (first entry)
XX
DE Tobacco PABFd repeated AT hook.
XX
KW Palindromic element binding factor; PABF; tobacco; cis-acting element;
KW transcription enhancer; heterologous promoter; AAT repeat element;
KW transcription factor; AT hook.
XX
OS Nicotiana tabacum.
XX
PN WO9749727-A1.
XX
PD 31-DEC-1997.
XX
PF 27-JUN-1997; 97WO-US011156.
XX
PR 27-JUN-1996; 96US-00669721.
XX
PA (SALK ) SALK INST BIOLOGICAL STUDIES.
XX
PI Lamb CJ, Doerner P, Laible G;
XX
DR WPI; 1998-077110/07.
XX
XX New isolated enhancer and transcription factor - used for increasing the
PT recombinant expression of proteins, particularly in plants, e.g. for
PT increasing production or providing pest resistance.
XX
PS Disclosure; Fig 7B; 65pp; English.
XX
AAW39790-W39796 are AT hook motifs used in the characterisation of a
CC novel tobacco palindromic element binding factor, (PABF). This PABF binds

```

CC to the sequence (AAATT)n where n at least 2. The (AAATT)n sequence has cis-  
CC acting, non-specific enhancer activity. It can be linked to a  
CC heterologous promoter operably linked with a gene to increase expression  
CC of the gene in a cell, particularly in plants. It can provide for  
CC increased expression of proteins such as nutritionally important  
CC proteins, growth promoting factors, proteins for early flowering in  
CC plants, proteins giving protection to the plant under certain  
CC environmental conditions, e.g. proteins conferring resistance to metals  
CC or other toxic substances, such as herbicides or pesticides, stress-  
CC related proteins which confer tolerance to temperature extremes, proteins  
CC conferring resistance to fungi, bacteria, viruses, insects and nematodes,  
CC proteins of specific commercial value, e.g. enzymes involved in metabolic  
CC pathways, such as EPS synthase. The PABF polypeptides act as  
CC transcription factor and bind to the (AAATT) repeat element to further  
CC boost the activity of the enhancer  
XX

SQ Sequence 11 AA;  
Query Match 100.0%; Score 27; DB 2; Length 11;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
|||||  
Db 4 GRRGR 8

RESULT 55  
AAW86177  
ID AAW86177 standard; peptide; 11 AA.

AC AAW86177;  
XX  
XX 04-MAR-1999 (first entry)

XX Peptide 31H used in a method for inhibiting wound contraction.  
DE Wound contraction; reduction; inhibition; tissue regeneration; scar;  
XX wound; joint motion; body deformation.  
KW Synthetic.

OS US5851994-A.  
XX  
XX 22-DEC-1998.

PF 06-JUN-1995; 95US-00473025.  
XX  
XX 28-APR-1994; 94US-00234979.

XX (LJOL-) LA JOLLA CANCER RES FOUND.  
PA Polarek J, Schreiber R;  
XX WPI; 1999-080478/07.

DR Inhibition of wound contraction - with peptide derivatives rich in basic  
XX amino acids.  
XX Claim 7; Col 24; 16pp; English.

XX The invention provides methods for reduction or inhibition of wound  
CC contraction that comprises administration of a peptide having more than 3  
CC consecutive basic amino acid residues. Alternatively, the peptide  
CC contains the amino acid sequence Arg-Gly-Asp and a basic amino acid  
CC sequence, or the peptide comprises 6-30 amino acids in which at least 4  
CC out of a sequence of 6 consecutive amino acids are basic amino acids. The  
CC method is used to allow normal tissue regeneration without excessive scar  
CC formation which, in the case of large wounds, can result in loss of joint  
CC motion or major body deformation. Sequences AAW86170 to AAW86183  
CC represent specifically claimed examples of peptides that can be used in  
CC the method of the invention  
XX

SQ Sequence 11 AA;  
Query Match 100.0%; Score 27; DB 2; Length 11;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
|||||  
Db 1 GRRGR 5

RESULT 56  
ABP54079  
ID ABP54079 standard; peptide; 11 AA.

AC ABP54079;  
XX  
XX 15-JAN-2003 (first entry)

XX Transport moiety cellular uptake peptide #3.  
DE Transporter; Spaced arginine moiety; vasotropic; neuroleptic; analgesic;  
XX antiparkinsonian; biologically active compound; biological membrane;  
KW epithelial tissue; endothelial tissue; ischaemia; neurotransmitter;  
XX schizophrenia; Parkinson's disease; pain; transport moiety.

OS Synthetic.  
XX  
XX Key Location/Qualifiers  
FT Modified-site 1 /label= ACP  
FT /note= "epsilon-aminocaproic acid; N-terminally modified  
FT with fluorescein (Fl)"  
FT Modified-site 11 /note= "C-terminally modified with CONH2"

XX WO200265986-A2.  
XX 29-AUG-2002.  
XX 14-FEB-2002; 2002WO-US004491.  
XX 16-FEB-2001; 2001US-00269627.  
XX (CELL-) CELLGATE INC.

XX Wender PA, Rothbard JB, Wright L, Kreider EL, Vandeusen CL;  
XX WPI; 2002-740700/80.  
XX Composition, useful for increasing the transport of a biologically active  
XX compound across a biological membrane, comprises a biologically active  
XX compound and a transport moiety.

XX Example 1; Page 23; 58pp; English.  
XX The present invention describes a composition (C) comprising a  
XX biologically active compound (A) and a transport moiety (B) of formula:  
XX (ZYZ)nZ (I), (ZYZ)nZ (II), (ZYZ)nZ (III) or (ZYZ)nZ (IV), where Z = L-  
XX arginine or D-arginine; Y = amino acid (not comprising amidino or  
XX guanidino moiety); and n = 2-10. Also described is a method for  
XX increasing the transport of a biologically active compound across a  
XX biological membrane involving administering (C). (C) has vasotropic,  
XX neuroleptic, antiparkinsonian and analgesic activities. (C) is used for  
XX increasing the transport of a biologically active compound across a  
XX biological membrane and across and into animal epithelial or endothelial  
XX tissues. (C) can be used for treating ischaemia and delivering  
XX neurotransmitters and other agents for treating schizophrenia,  
XX Parkinson's disease and pain. The transport of the biologically active  
XX compound across the biological membrane is increased relative to the  
XX transport of the biologically active compound in the absence of the  
XX transport moiety. The present sequence represents a transport moiety  
XX cellular uptake peptide, which is used in an example from the present

CC invention  
 XX Sequence 11 AA;  
 SQ Query Match 100.0%; Score 27; DB 5; Length 11;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRGR 5  
 DB 3 GRGR 7

RESULT 57  
 ADC19823  
 ID ADC19823 standard; peptide; 11 AA.  
 XX  
 AC ADC19823;  
 DT 18-DEC-2003 (first entry)  
 XX  
 DE Fluorescently labelled spaced arginine transport peptide #3.  
 XX  
 KW Cellular membrane transport peptide; epithelial tissue;  
 KW endothelial tissue; drugs transport; stratum corneum; antibacterial;  
 KW antifungal; antiviral; antiproliferative; immunosuppressive; vitamin;  
 KW analgesic; hormone.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Modified-site 1  
 FT /label  
 FT /note= "xaa is fluorescently labelled epsilon-  
 aminocaproic acid"  
 FT Modified-site 11  
 FT /label= OTHER  
 FT /note= "Arg is covalently bound to a CONH2 group"  
 FT  
 XX US2003032593-A1.  
 XX  
 PD 13-FEB-2003.  
 XX  
 XX 14-FEB-2002; 2002US-00078247.  
 XX  
 XX 16-FEB-2001; 2001US-0269627P.  
 XX  
 XX (CELL-) CELLGATE INC.  
 XX  
 XX Wender PA, Rothbard JB, Wright L, Kreider EL, Vandeusen CJ;  
 XX WPI; 2003-786846/74.  
 XX  
 PT Composition used for increasing transport of biologically active compound  
 PT across biological membrane comprises biologically active compound and  
 PT transport group.  
 PS Example 1; Page 9; 33pp; English.  
 XX  
 CC The invention relates to a composition comprising a biologically active  
 CC compound and a transport group. The transport group comprises a spaced  
 CC poly-Arginine based peptide of formula given in the specification. The  
 CC spaced poly-Arginine based peptide acts as a cellular membrane transport  
 CC signal and effects transport of the biologically active compound across  
 CC the membrane. The conjugate is also useful in therapeutic, prophylactic  
 CC and diagnostic applications. The composition improves the transport of  
 CC biologically active compounds across the biological membrane and into  
 CC animal epithelial or endothelial tissues. The arginine residue of the  
 CC conjugate provides an enhanced transport of drugs and are a part of the  
 CC polypeptide that provides suitable spacing between arginine residues. The  
 CC transport groups deliver an agent across the stratum corneum, which  
 CC previously had been a nearly impenetrable barrier to drug delivery. The  
 CC ability of the conjugate to obtain penetration of skin layers improves

CC the efficacy of compounds such as antibacterials, antifungals,  
 CC antivirals, antiproliferatives, immunosuppressives, vitamins, analgesics  
 CC and hormones. The present sequence is a Fluorescently labelled spaced  
 CC arginine transport peptide of the invention.  
 XX Sequence 11 AA;  
 SQ Query Match 100.0%; Score 27; DB 7; Length 11;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GRGR 5  
 DB 3 GRGR 7

RESULT 58  
 ADG44414  
 ID ADG44414 standard; peptide; 11 AA.  
 XX  
 AC ADG44414;  
 XX  
 DT 26-FEB-2004 (first entry)  
 XX  
 DE Anti-pathogen related amino acid sequence SEQ ID NO:213.  
 XX  
 KW pathogen infection; chimeric molecule; pathogen-detection domain;  
 KW pathogen-induced product-detection domain; effector domain;  
 KW antimicrobial; gene therapy.  
 XX  
 OS Synthetic.  
 XX  
 PN WO2003066003-A2.  
 XX  
 PD 14-AUG-2003.  
 XX  
 XX 07-FEB-2003; 2003WO-US003978.  
 XX  
 XX 07-FEB-2002; 2002US-0355022P.  
 PR 07-FEB-2002; 2002US-0355359P.  
 PR 10-DEC-2002; 2002US-0432386P.  
 XX  
 PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
 XX  
 PI Rider TH;  
 XX  
 DR WPI; 2003-731477/59.  
 XX  
 PT Treating or preventing a pathogen infection in a cell by administering to  
 PT the cell chimeric molecules having a pathogen-detection domain or  
 PT pathogen-induced product-detection domain and at least one effector  
 PT domain.  
 XX  
 PS Example 7; SEQ ID NO 213; 348pp; English.  
 XX  
 CC The present invention describes a method for treating or preventing a  
 CC pathogen infection in a cell. The method comprises administering to the  
 CC cell chimeric molecules (I) having at least one pathogen-detection domain  
 CC or pathogen-induced product-detection domain and at least one effector  
 CC domain. The pathogen-detection domain or pathogen-induced product-  
 CC detection domain is not naturally bound to the effector domain. In the  
 CC presence of a pathogen or pathogen-induced product in the cell, the  
 CC chimeric molecules bind to the pathogen and activate the effector domain.  
 CC Also described: (1) treating or preventing the spread of a pathogen  
 CC infection in an organism; (2) a chimeric molecule (I) having at least one  
 CC pathogen-detection domain or pathogen-induced product-detection domain  
 CC and at least one effector domain; (3) an agent having at least one  
 CC pathogen-interacting or pathogen-induced product-interacting molecular  
 CC structure and at least one effector-mediated molecular structure; and  
 CC (4) an assay for detecting a pathogen infection in a cell or organism.  
 CC (I) has antimicrobial activity, and can be used in gene therapy. The  
 CC method is useful for treating or preventing a pathogen infection in a  
 CC cell or organism. The present sequence is used in the exemplification of





RESULT 61  
AAW65568  
ID AAW65568 standard; peptide; 12 AA.  
XX AC AAW65568;  
XX AC AAW65568;  
XX AC AAW65568;  
DT 27-AUG-2003 (revised)  
DT 15-OCT-1998 (first entry)  
XX DT  
DE Epstein-Barr virus derived peptide #19.  
XX  
KW Epstein-Barr virus; diagnostic test; autoimmune disease; vaccine;  
KW infection; antibody; screening; genetic marker.  
XX  
OS Synthetic.  
OS Human herpesvirus 4.  
XX  
XX  
PN WO9830586-A2.  
XX  
XX  
PD 16-JUL-1998.  
XX  
XX  
PF 13-JAN-1998; 98WO-US000342.  
XX  
PR 13-JAN-1997; 97US-00781296.  
XX  
XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
XX  
PI Harley JB, James JA;  
XX  
XX WPI; 1998-399062/34.  
DR  
XX  
XX Use of Epstein-Barr virus or component(s) - for developing product(s)  
PT which can be used for preventing, diagnosing, treating or determining  
PT risk of developing autoimmune disease.  
PT  
XX  
PS Claim 8; Page 64; 81pp; English.  
XX  
XX The invention relates to a vaccine for alleviating or preventing  
CC autoimmune disorders induced by infection with Epstein-Barr virus (EBV).  
CC It comprises EBV or a component in a carrier for administration of the  
CC virus or viral component to alleviate or prevent the autoimmune disorder.  
CC Also claimed are: (1) a diagnostic test kit comprising: (a) reagents  
CC which can be used to detect levels of antibodies to EBV, indicators of  
CC EBV infection of cells, or levels of EBV DNA or protein in a patient; (b)  
CC control samples from individuals not at risk of developing an autoimmune  
CC disease; and (c) a device for determining the differences in levels of a  
CC patient and control samples to distinguish individuals at higher risk of  
CC developing an autoimmune disease from those at lower risk of developing  
CC an autoimmune disease; and (2) a method for screening for genetic markers  
CC or risk factors for development of autoimmune disorders induced by  
CC infection with EBV comprising comparing the responses of different  
CC strains of the same species of an animal vaccinated with EBV or a  
CC component to induce an autoimmune response in at least one of the strains  
CC and comparing the differences in the genetics of the different strains to  
CC identify potential genetic markers or risk factors. The methods can be  
CC used for the prevention, diagnosis, and treatment of autoimmune diseases  
CC having EBV as an etiological agent. The autoimmune diseases may be e.g.  
CC systemic lupus erythematosus, Sjogren's syndrome, rheumatoid arthritis,  
CC juvenile onset diabetes mellitus, Wegener's granulomatosis, etc. The  
CC present sequence represents a peptide derived from Epstein-Barr virus.  
CC Reagents are used to detect antibodies to this peptide in a specifically  
CC claimed diagnostic test. (Updated on 27-AUG-2003 to correct OS field.)  
XX  
SQ Sequence 12 AA;

Query Match 100.0%; Score 27; DB 2; Length 12;  
Best Local Similarity 100.0%; Pred. NO. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
|||  
6 GRRR 10

Db

RESULT 62  
AAE09159  
ID AAE09159 standard; peptide; 12 AA.  
XX AC AAE09159;  
XX AC AAE09159;  
XX AC AAE09159;  
DT 11-SEP-2003 (revised)  
DT 15-NOV-2001 (first entry)  
XX DT  
DE Epstein-Barr virus (EBV) peptide #4 used in the invention.  
XX  
KW Vaccine; therapy; autoimmune disorder; Epstein-Barr virus; EBV; anaemia;  
KW polyomyelitis; systemic lupus erythematosus; SLE; rheumatoid arthritis;  
KW Sjogren's syndrome; diabetes mellitus; adenitis; multiple sclerosis;  
KW demyelinating disease; Hashimoto's thyroiditis; autoimmune infertility;  
KW hypoparathyroidism; primary biliary cirrhosis; ankylosing spondylitis;  
KW inflammatory bowel disease; Addison's disease; thyroiditis; filariasis;  
KW Graves' disease; ulcerative colitis; dermatomyositis; myasthenia gravis;  
KW Crohn's disease; CREST syndrome; autoimmune cystitis; glomerulonephritis;  
KW polyarteritis nodosa; hepatitis; atopic rhinitis; Goodpasture's syndrome;  
KW sarcoidosis; rheumatic fever; anti-phospholipid syndrome; farmer's lung;  
KW Cushing's syndrome; bird-fancier's lung; alveolitis; erythema nodosum;  
KW pyoderma gangrenosum; fibromyalgia; Kawasaki's disease; cardiomyopathy;  
KW Sampter's syndrome; asthma; polymyalgia rheumatica; psoriasis; arteritis;  
KW erythroblastosis foetalis; cyclitis; IGA nephropathy; Hodgkin's lymphoma;  
KW renal cell carcinoma; eosinophilia; immunosuppressive; ophthalmological;  
KW thrombotic; neuroprotective; cytostatic; nephrotropic; antiallergic;  
KW dengue; antitumor; vasotropic; antipyretic; hepatotropic.  
XX  
OS Human herpesvirus 4.  
XX  
XX WO200158481-A2.  
XX  
XX 16-AUG-2001.  
XX  
XX 09-FEB-2001; 2001WO-US004191.  
XX  
XX 09-FEB-2000; 2000US-00500904.  
XX (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
XX  
PI Harley JB, James JA, Kaufman KM;  
XX  
XX WPI; 2001-522437/57.  
XX  
XX Novel vaccine for alleviating or preventing autoimmune disorders induced  
PT Epstein-Barr virus (EBV) infection e.g. systemic lupus erythematosus,  
PT juvenile onset diabetes mellitus, comprises EBV virus or its component.  
XX  
PS Claim 8; Page 61; 114pp; English.

The present invention relates to a vaccine for alleviating or preventing  
CC autoimmune disorders induced by infection with Epstein-Barr virus (EBV),  
CC comprising EBV or its component in a carrier. The vaccine is useful for  
CC preventing or alleviating autoimmune disorders induced by EBV, e.g.  
CC systemic lupus erythematosus (SLE), Sjogren's syndrome, juvenile onset  
CC diabetes mellitus, rheumatoid arthritis, Wegener's granulomatosis,  
CC inflammatory bowel disease, polymyositis, dermatomyositis, multiple  
CC endocrine failure, Schmidt's syndrome, autoimmune uveitis, Addison's  
CC disease, adenitis, primary biliary cirrhosis, Graves' disease,  
CC thyroiditis, Hashimoto's thyroiditis, autoimmune thyroid disease,  
CC pernicious and haemolytic anaemia, lupoid hepatitis, demyelinating  
CC disease, multiple sclerosis, subacute cutaneous lupus erythematosus,  
CC hypoparathyroidism, Dressler's syndrome, myasthenia gravis, autoimmune  
CC idiopathic thrombocytopenic purpura, autoimmune pemphigus vulgaris,  
CC pemphigus, bullous pemphigoid, dermatitis herpetiformis, alopecia areata,  
CC autoimmune cystitis, pemphigoid, scleroderma, progressive systemic  
CC sclerosis, CREST syndrome (calcinosis, Raynaud's oesophageal dysmotility,  
CC sclerodactyly and telangiectasia), adult onset diabetes mellitus (Type II  
CC diabetes), male or female autoimmune infertility, ankylosing spondylitis,  
CC ulcerative colitis, Crohn's disease, mixed connective tissue disease,

CC polyarteritis nodosa, systemic necrotising vasculitis,  
CC glomerulonephritis, atopic dermatitis, atopic rhinitis, Goodpasture's  
CC syndrome, Chagas' disease, sarcoidosis, rheumatic fever, asthma,  
CC recurrent abortion, anti-phospholipid syndrome, farmer's lung, erythema  
CC multiforme, postcardotomy syndrome, Cushing's syndrome, autoimmune  
CC chronic active hepatitis, bird-fancier's lung, allergic  
CC encephalomyelitis, toxic necrodermal lysis, alopecia, Alport's syndrome,  
CC allergic alveolitis, fibrosing alveolitis, interstitial lung disease,  
CC erythema nodosum, pyoderma gangrenosum, transfusion reaction, chronic  
CC fatigue syndrome, fibromyalgia, Takayasu's arteritis, Kawasaki's disease,  
CC polymyalgia rheumatica, temporal arteritis, giant cell arteritis, dengue,  
CC Sandler's syndrome (triaditis, nasal polyps, eosinophilia) and Behcet's  
CC disease, Caplan's syndrome, encephalomyositis, erythema elevatum et  
CC diutinum, psoriasis, erythroblastosis foetalis, Shulman's syndrome, Iga  
CC nephropathy, Felty's syndrome, fascitis with eosinophilia, filariasis,  
CC chronic cyclitis, heterochromic cyclitis, Fuch's cyclitis, Hodgkin's and  
CC non-Hodgkin's lymphoma, cardiomyopathy, Henoch-Schonlein purpura, post  
CC vaccination syndromes, renal cell carcinoma, Eaton-Lambert syndrome or  
CC relapsing polycondritis. The present sequence is a EBV peptide used in  
CC the invention. (Updated on 11-SEP-2003 to standardise OS field)

XX  
SQ Sequence 12 AA;  
Query Match 100.0%; Score 27; DB 4; Length 12;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
Db |||||  
6 GRRGR 10

RESULT 63  
ABB74766  
ID ABB74766 standard; peptide; 12 AA.  
XX  
AC ABB74766;  
XX  
DT 18-APR-2002 (first entry)  
XX  
DE Nuclear protein nuclear localisation signal peptide SEQ ID NO:530.  
XX  
KW Fusogenic; nuclear localisation signal; NLS; encapsulation; lipogene;  
XX liposome; micelle; karyophilic; cytotatic; antitumour; solid tumour;  
XX peptide-lipid-polynucleotide complex; neoplastic disease; gene therapy;  
XX breast carcinoma; prostate carcinoma.  
XX  
OS Saccharomyces cerevisiae.  
XX  
PN WO200193836-A2.  
XX  
PD 13-DEC-2001.  
XX  
PF 08-JUN-2001; 2001WO-US018657.  
XX  
PR 09-JUN-2000; 2000US-0210925P.  
XX  
PA (BOUL/) BOULIKAS T.  
XX  
PI Boulikas T;  
XX  
DR WPI; 2002-164295/21.  
XX  
PT Encapsulation of plasmid DNA (Lipogenes) and therapeutic agents with  
PT nuclear localization signal/fusogenic peptide conjugates into targeted  
PT liposome complexes.  
XX  
PS Claim 14; Page 83; 107pp; English.  
XX  
CC The present invention describes a method for producing micelles with  
CC entrapped therapeutic agents. The method comprises: (i) combining  
CC negatively charged agent with a cationic lipid in a ratio where 30-90 %  
CC of the negatively charged atoms are neutralised by positive charges on

CC lipid molecules to form an electrostatic micelle complex in 20-80 %  
CC ethanol; and (2) combining the micelle complex of (a) with fusogenic-  
CC karyophilic peptide conjugates in a 0.0-0.3 ratio, therefore producing  
CC micelles with entrapped therapeutic agents. Also described is a method  
CC for delivering a therapeutic agent in vivo, comprising the administration  
CC of the micelle. ABB74256 to ABB74858 represent specifically claimed  
CC nuclear localisation signal (NLS) peptides for use in the method as the  
CC fusogenic-karyophilic peptides. The micelles produced can have cytostatic  
CC and antitumour activities. The peptide-lipid-polynucleotide complexes  
CC produced are useful for inhibiting the progression of neoplastic  
CC diseases. The invention relates to the field of gene therapy and is  
CC directed toward methods for producing peptide-lipid-polynucleotide  
CC complexes suitable for delivery of polynucleotides. The encapsulated  
CC molecules display therapeutic efficacy in eradicating solid tumours  
CC including but not limited to breast carcinoma or prostate carcinoma.  
CC ABB74235 to ABB74255 are used in the exemplification of the present  
CC invention

XX  
SQ Sequence 12 AA;  
Query Match 100.0%; Score 27; DB 5; Length 12;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
Db |||||  
6 GRRGR 10

RESULT 64  
ABU07643  
ID ABU07643 standard; peptide; 12 AA.  
XX  
AC ABU07643;  
XX  
DT 23-OCT-2003 (revised)  
DT 10-MAY-2003 (first entry)  
XX  
DE Epstein-Barr virus nuclear antigen peptide #16.  
XX  
KW EBV; viral; Epstein-Barr virus nuclear antigen; vaccine;  
KW autoimmune disorder; Epstein-Barr virus infection; diabetes mellitus;  
KW arthritis; multiple sclerosis; dermatitis; psoriasis; asthma; anaemia;  
KW allergy.  
XX  
OS Human herpesvirus 4.  
XX  
PN US2002164355-A1.  
XX  
PD 07-NOV-2002.  
XX  
PF 24-OCT-2001; 2001US-00012756.  
XX  
PR 30-NOV-1993; 93US-00160604.  
PR 16-MAY-1996; 96US-0019053P.  
PR 13-JAN-1997; 97US-00781296.  
XX  
PA (HARL/) HARLEY J B.  
PA (JAME/) JAMES J A.  
XX  
PI Harley JB, James JA;  
XX  
DR WPI; 2003-298686/29.  
XX  
PT New vaccine preventing or alleviating autoimmune disorders induced by the  
PT Epstein-Barr virus, such as diabetes mellitus, rheumatoid arthritis,  
PT multiple sclerosis, systemic lupus erythematosus, atopic dermatitis and  
PT psoriasis.  
XX  
PS Claim 28; Page 28; 41pp; English.  
XX  
CC The invention relates to a vaccine for alleviating or preventing  
CC autoimmune disorders induced by infection with Epstein-Barr virus,

CC comprising an Epstein-Barr virus or a component in a carrier for  
 CC administration to alleviate or prevent the autoimmune disorders. The  
 CC methods and compositions of the present invention are useful for  
 CC diagnosing, preventing, treating and/or alleviating autoimmune disorders,  
 CC such as diabetes mellitus, rheumatoid arthritis, juvenile rheumatoid  
 CC arthritis, osteoarthritis, psoriatic arthritis, multiple sclerosis,  
 CC encephalomyelitis, myasthenia gravis, systemic lupus erythematosus,  
 CC autoimmune thyroiditis, atopic dermatitis, eczematous dermatitis,  
 CC psoriasis, Sjogren's Syndrome, Crohn's disease, aphthous ulcer, iritis,  
 CC conjunctivitis, keratoconjunctivitis, ulcerative colitis, asthma,  
 CC allergic asthma, cutaneous lupus erythematosus, scleroderma, vaginitis,  
 CC proctitis, drug eruptions, leprosy reversal reactions, erythema  
 CC nodosum, erythema multiforme, erythema nodosum, erythema  
 CC necrotising haemorrhagic encephalopathy, idiopathic bilateral progressive  
 CC sensorineural hearing loss, aplastic anaemia, pure red cell anaemia,  
 CC idiopathic thrombocytopenia, polychondritis, Wegener's granulomatosis,  
 CC chronic active hepatitis, Stevens-Johnson syndrome, idiopathic sprue,  
 CC lichen planus, Graves' disease, sarcoidosis, primary biliary cirrhosis,  
 CC uveitis posterior, interstitial lung fibrosis, graft-versus-host disease,  
 CC and allergy. The present sequence represents an Epstein-Barr virus  
 CC nuclear antigen peptide used in the method of the invention. (Updated on  
 CC 23-OCT-2003 to standardise OS field)  
 XX  
 SQ Sequence 12 AA;

Query Match 100.0%; Score 27; DB 6; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
 DB 6 GRGR 10

# RESULT 65

AD017059

ID AD017059 standard; peptide; 12 AA.  
 AC AD017059;  
 XX  
 DT 29-JUL-2004 (first entry)  
 XX  
 DE Epstein-Barr virus nuclear antigen 1, antigenic peptide #7.  
 XX  
 KW vaccine; autoimmune disorder; Epstein-Barr virus; EBV;  
 KW systemic lupus erythematosus; Sjogren's syndrome; rheumatoid arthritis;  
 KW juvenile onset diabetes mellitus; Wegener's granulomatosis;  
 KW inflammatory bowel disease; Epstein-Barr nuclear antigen 1.  
 XX  
 OS Human herpesvirus 4.  
 XX  
 PN US2004086522-A1.  
 XX  
 PD 06-MAY-2004.  
 XX  
 PF 27-JUN-2003; 2003US-00607918.  
 XX  
 PR 10-NOV-1993; 93US-00160604.  
 PR 16-MAY-1996; 96US-0019053P.  
 PR 13-JAN-1997; 97US-00781296.  
 PR 24-OCT-2001; 2001US-00012756.  
 XX  
 PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.  
 XX  
 PI Harley JB, James JA;  
 XX  
 DR WPI; 2004-356164/33.  
 XX  
 CC New vaccine for alleviating or preventing autoimmune disorders induced by  
 PT Epstein-Barr virus (EBV) infection, e.g. systemic lupus erythematosus,  
 PT comprises EBV or its component in a pharmaceutical carrier.  
 XX  
 PS Claim 8; Page 17; 30pp; English.

XX The invention relates to a new vaccine for alleviating or preventing  
 CC autoimmune disorders induced by infection with Epstein-Barr virus (EBV).  
 CC The vaccine comprises EBV or its component in a pharmaceutical carrier  
 CC for administration of the virus or viral component in an amount and mode  
 CC of administration to alleviate or prevent the autoimmune disorders. The  
 CC composition and methods are useful for diagnosing, preventing or treating  
 CC autoimmune diseases caused by Epstein-Barr virus, such as systemic lupus  
 CC erythematosus, Sjogren's syndrome, rheumatoid arthritis, juvenile onset  
 CC diabetes mellitus, Wegener's granulomatosis or inflammatory bowel  
 CC disease. These may also be used in screening of therapeutics for  
 CC prevention or alleviation of autoimmune disorders induced by EBV  
 CC infection. The present sequence represents an Epstein-Barr virus nuclear  
 CC antigen 1, antigenic peptide used to make the vaccine of the invention.  
 XX  
 SQ Sequence 12 AA;

Query Match 100.0%; Score 27; DB 8; Length 12;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
 DB 6 GRGR 10

# RESULT 66

AAR98450

ID AAR98450 standard; peptide; 13 AA.

AC AAR98450;

DT 12-FEB-1997 (first entry)

DE Peptide with cell attachment promoting activity.

XX  
 KW Hyaluronate; treatment; wound healing; burns; skin graft; ulcer;  
 KW tissue regeneration; matrix; matrices; fibronectin; cancer; osteoporosis;  
 KW thrombosis.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Modified-site 13 /note= "Amidated"  
 XX  
 PN WO9620002-A1.  
 XX  
 PD 04-JUL-1996.  
 XX  
 PF 21-DEC-1995; 95WO-US016959.  
 XX  
 PR 23-DEC-1994; 94US-00363213.  
 XX  
 PA (LJOL-) LA JOLLA CANCER RES FOUND.  
 XX  
 PI Dickerson KT, Glass JR, Liu L, Polarek JW, Craig WS, Mullen DG;  
 PI Cheng S;  
 XX  
 DR WPI; 1996-321641/32.  
 XX  
 PT Crosslinked hyaluronate-RGD peptide conjugates - prepd. by epoxide,  
 PT sodium periodate or treayl chloride methods, provide temporary matrix for  
 PT wound healing and tissue regeneration.  
 XX  
 PS Claim 42; Page 35; 48pp; English.

XX Compositions comprising a cross linked hyaluronate polymer and a peptide  
 CC having cell attachment promoting activity can be used for treating wounds  
 CC such as severe burns, skin graft donor sites, decubitus ulcers, diabetic  
 CC ulcers, surgical incisions and keloid- forming wounds. They can also be  
 CC used for inducing tissue regeneration. The conjugate acts as a temporary  
 CC replacement matrix that encourages cell migration into the wound and

CC speeds healing. As the wound heals, the conjugate is slowly broken down  
CC by the migrating cells and is replaced by a natural matrix. The peptides  
CC contain the amino acid sequence Y-Gly-Asp, where Y is Arg or D-Arg, at  
CC least two more amino acids selected from (D-)Arg, (D-)Lys, (D-)Orn and (D  
CC -)HomoArg. The peptides can be used to competitively inhibit the binding  
CC of cells to RGD containing adhesive proteins such as fibronectin for the  
CC treatment of e.g cancer, osteoporosis or thrombosis. Peptides are  
CC described in AAR98433-36, AAR98438-43 and AAR98445-53  
XX  
SQ Sequence 13 AA;

Query Match 100.0%; Score 27; DB 2; Length 13;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
Db 1 GRRGR 5  
RESULT 67  
AAW93300  
ID AAW93300 standard; peptide; 13 AA.  
AC AAW93300;  
XX  
DT 27-MAY-1999 (first entry)  
XX  
DE C1q peptide mimic #6.  
XX  
KW C1q; active centre; adsorbent; binding; immune complex; amyloid; virus;  
KW pathogen; endotoxin; C-reactive protein; cardiolipid; fibronectin; mimic;  
KW fibrinogen; solid carrier; selective; blood; plasma; serum.  
XX  
OS Synthetic.  
XX  
PN DE19735902-A1.  
XX  
PD 25-FEB-1999.  
XX  
PF 19-AUG-1997; 97DE-01035902.  
XX  
PR 19-AUG-1997; 97DE-01035902.  
XX  
PA (IMTE-) IMTEC IMMUNDIAGNOSTIKA GMBH.  
XX

PI Schoessler W, Hiepe F, Hentschel C, Pfeueller B;  
DR WPI; 1999-154848/14.  
XX  
PT New selective adsorbent carrying peptide(s) that mimic complement C1q -  
PT useful for removing immune complexes, viruses etc. from blood, plasma and  
PT serum.  
XX  
PS Claim 7; Col 1; 4pp; German.  
XX  
CC This invention describes a novel selective adsorbent for binding immune  
CC complexes, viruses, other pathogens, endotoxins, C-reactive proteins,  
CC amyloid, DNA, cardiolipids, fibronectin, fibrinogen and other biological  
CC materials. The adsorbent comprises a solid carrier to which are bound  
CC synthetic peptides which mimic the active center of protein C1q. The  
CC adsorbent is useful for eliminating specified materials from blood,  
CC plasma and serum, for medical or pharmaceutical applications. The  
CC adsorbent can be produced with consistent quality from readily available  
CC synthetic peptides. The peptides of the invention bind the specified  
CC materials with high affinity and selectivity  
XX  
SQ Sequence 13 AA;

Query Match 100.0%; Score 27; DB 2; Length 13;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
Db 1 GRRGR 5  
RESULT 67  
AAW93300  
ID AAW93300 standard; peptide; 13 AA.  
AC AAW93300;  
XX  
DT 27-MAY-1999 (first entry)  
XX  
DE C1q peptide mimic #6.  
XX  
KW C1q; active centre; adsorbent; binding; immune complex; amyloid; virus;  
KW pathogen; endotoxin; C-reactive protein; cardiolipid; fibronectin; mimic;  
KW fibrinogen; solid carrier; selective; blood; plasma; serum.  
XX  
OS Synthetic.  
XX  
PN DE19735902-A1.  
XX  
PD 25-FEB-1999.  
XX  
PF 19-AUG-1997; 97DE-01035902.  
XX  
PR 19-AUG-1997; 97DE-01035902.  
XX  
PA (IMTE-) IMTEC IMMUNDIAGNOSTIKA GMBH.  
XX

PI Schoessler W, Hiepe F, Hentschel C, Pfeueller B;  
DR WPI; 1999-154848/14.  
XX  
PT New selective adsorbent carrying peptide(s) that mimic complement C1q -  
PT useful for removing immune complexes, viruses etc. from blood, plasma and  
PT serum.  
XX  
PS Claim 7; Col 1; 4pp; German.  
XX  
CC This invention describes a novel selective adsorbent for binding immune  
CC complexes, viruses, other pathogens, endotoxins, C-reactive proteins,  
CC amyloid, DNA, cardiolipids, fibronectin, fibrinogen and other biological  
CC materials. The adsorbent comprises a solid carrier to which are bound  
CC synthetic peptides which mimic the active center of protein C1q. The  
CC adsorbent is useful for eliminating specified materials from blood,  
CC plasma and serum, for medical or pharmaceutical applications. The  
CC adsorbent can be produced with consistent quality from readily available  
CC synthetic peptides. The peptides of the invention bind the specified  
CC materials with high affinity and selectivity  
XX  
SQ Sequence 13 AA;

Query Match 100.0%; Score 27; DB 2; Length 13;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
Db 5 GRRGR 9  
RESULT 68  
AAB36917  
ID AAB36917 standard; peptide; 13 AA.  
XX  
AC AAB36917;  
XX  
DT 26-FEB-2001 (first entry)  
XX  
DE Peptide C1qA (14-26).  
XX  
KW Gene therapy; anionic transfer.  
XX  
OS Homo sapiens.  
XX  
PN EP1052287-A2.  
XX  
PD 15-NOV-2000.  
XX  
PF 08-MAY-2000; 2000EP-00401284.  
XX  
PR 10-MAY-1999; 99EP-00401155.  
PR 03-MAR-2000; 2000US-0187217P.  
XX  
PA (TRGE ) TRANSGENE SA.  
XX  
PI Jacobs E;  
XX  
DR WPI; 2001-042330/06.  
XX  
PT New complex for transferring an anionic substance such as a  
PT polynucleotide into a cell comprising all or part of the C1 complement  
PT factor amino acid sequence, useful to introduce polynucleotides into  
PT cells in gene therapy.  
XX  
PS Example 4; Page 13; 19pp; English.  
XX  
CC The present invention relates to a complex for transferring an anionic  
CC substance e.g. a polynucleotide into a cell. The complexes are useful for  
CC transferring an anionic substance into a cell, especially a  
CC polynucleotide for gene therapy  
XX  
SQ Sequence 13 AA;  
Query Match 100.0%; Score 27; DB 4; Length 13;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
Db 5 GRRGR 9  
RESULT 69  
AAB36910  
ID AAB36910 standard; peptide; 13 AA.  
XX  
AC AAB36910;  
XX  
DT 26-FEB-2001 (first entry)  
XX  
DE C1qB peptide.  
XX  
KW Gene therapy; anionic transfer.  
XX  
OS Synthetic.  
XX  
PN EP1052288-A1.  
XX

PD 15-NOV-2000.  
 XX  
 PF 10-MAY-1999; 99EP-00401155.  
 PR 10-MAY-1999; 99EP-00401155.  
 XX (TRGE ) TRANSGENE SA.  
 PA  
 XX Jacob E;  
 PI  
 XX WPI; 2001-042331/06.  
 DR  
 XX  
 CC New complex for transferring an anionic substance such as a  
 PT polynucleotide into a cell comprising all or part of the C1 complement  
 PT factor amino acid sequence, useful to introduce polynucleotides into  
 PT cells in gene therapy.  
 XX  
 XX Claim 4; Page 14; 23pp; English.  
 PS  
 XX  
 CC The present invention relates to a complex for transferring an anionic  
 CC substance e.g. a polynucleotide into a cell. The complexes are useful for  
 CC transferring an anionic substance into a cell, especially a DNA for gene  
 CC therapy  
 CC  
 XX Sequence 13 AA;  
 SQ  
 Query Match 100.0%; Score 27; DB 4; Length 13;  
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GRGR 5  
 Db 5 GRGR 9  
 RESULT 70  
 AD238438  
 ID AD238438 standard; peptide; 13 AA.  
 XX  
 AC AD238438;  
 XX  
 DT 30-JUN-2005 (first entry)  
 XX  
 DE Human kinase substrate peptide SEQ ID 1368.  
 XX  
 KW Kinase; substrate; drug delivery; cancer; restenosis; osteoporosis;  
 KW rheumatoid arthritis; asthma; psoriasis; inflammatory bowel disease;  
 KW systemic lupus erythematosus; multiple sclerosis; transplant rejection;  
 KW neoplasm; cytostatic; vasotropic; cardiovascular disease; osteopathic;  
 KW degeneration; endocrine disease; musculoskeletal disease;  
 KW antiinflammatory; inflammation; autoimmune disease; immunosuppressive;  
 KW immune disorder; antiarthritic; antirheumatic; antiasthmatic;  
 KW respiratory disease; antipsoriatic; dermatological disease;  
 KW gastrointestinal-gen.; gastrointestinal disease; neuroprotective;  
 KW neurological disease; dermatological; dermatological disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2005035003-A2.  
 XX  
 XX 21-APR-2005.  
 XX  
 XX 22-SEP-2004; 2004WO-US031148.  
 XX  
 PR 22-SEP-2003; 2003US-0505325P.  
 PR 04-MAY-2004; 2004US-0568340P.  
 PR 22-JUN-2004; 2004US-0581835P.  
 XX  
 XX (DIHE-) DIHEDRON CORP.  
 PA  
 XX Ballatore C, Castellino AJ, Desharnais J, Guo Z, Li Q, Newman MJ;  
 PI Sun C;  
 XX

DR  
 XX  
 PT New conjugate used for treating aberrant cellular activation, migration,  
 PT proliferation or survival condition such as cancer, comprises drug and  
 PT substrate for protein or lipid kinase linked to protein, optionally by  
 PT non-releasable linker.  
 XX  
 PS Disclosure; SEQ ID NO 1368; 407pp; English.  
 XX  
 CC The invention relates to a conjugate (I) comprising a drug and a  
 CC substrate for a protein kinase or a lipid kinase non-releasably linked to  
 CC it, optionally by a non-releasable linker. Also included are preparing a  
 CC pacitaxel C10 carbanate of formula (8a) (which comprises reacting a  
 CC pacitaxel compound of formula (5a) with a carbodiimide compound and  
 CC reacting the obtained compound of formula (6a) with an amine of formula  
 CC X, all formulae given in the specification), a pharmaceutical composition  
 CC (PCI) comprising (I) and a carrier, an article (comprising packaging  
 CC material, (I) or its derivatives, for treatment/prevention/amelioration  
 CC of one or more symptoms (associated with aberrant cellular activation,  
 CC migration, proliferation or survival (ACAMPS) and a label that indicates  
 CC that (I) is used for treatment, prevention or amelioration of one or more  
 CC symptoms associated with ACAMPS), and a peptide comprising an kinase  
 CC substrate peptide. The conjugate has improved cytotoxic selectivity index  
 CC as compared to an unconjugated drug. The conjugate is useful for treating  
 CC conditions caused by ACAMPS characterized by undesirable or aberrant  
 CC activation, migration, proliferation or survival of tumor cells,  
 CC endothelial cells, B cells, T cells, macrophages, neutrophils,  
 CC eosinophils, basophils, monocytes, platelets, fibroblasts, other  
 CC connective tissue cells, osteoblasts, osteoclasts and progenitors of  
 CC these cell types. The ACAMPS condition is a cancer, coronary restenosis,  
 CC osteoporosis, chronic inflammation or autoimmunity disease. The  
 CC autoimmune disease is rheumatoid arthritis, asthma, psoriasis,  
 CC inflammatory bowel disease, systemic lupus erythematosus, systemic  
 CC dermatomyositis, inflammatory ophthalmic diseases, autoimmune hematologic  
 CC disorders, multiple sclerosis, vasculitis, idiopathic nephrotic syndrome,  
 CC transplant rejection or graft versus host disease. The cancer is non-  
 CC small cell lung cancer, head squamous cancer, neck squamous cancer,  
 CC colorectal cancer, prostate cancer, breast cancer, acute lymphocytic  
 CC leukemia, adult acute myeloid leukemia, adult non-Hodgkin's lymphoma,  
 CC brain tumor, cervical cancer, childhood cancer, childhood sarcoma,  
 CC chronic lymphocytic leukemia, chronic myeloid leukemia, esophageal  
 CC cancer, hairy cell leukemia, kidney cancer, liver cancer, multiple  
 CC myeloma, neuroblastoma, oral cancer, pancreatic cancer, primary central  
 CC nervous system lymphoma, skin cancer or small-cell lung cancer. The  
 CC cancer is brain stem glioma, cerebellar astrocytoma, cerebral  
 CC astrocytoma, ependymoma, Ewing's sarcoma, germ cell tumor, Hodgkin's  
 CC disease, acute myelogenous leukemia, acute lymphoblastic leukemia, liver  
 CC cancer, medulloblastoma, neuroblastoma, non-Hodgkin's lymphoma,  
 CC osteosarcoma, malignant fibrous histiocytoma of bone, retinoblastoma,  
 CC rhabdomyo sarcoma, soft tissue sarcoma, supratentorial primitive  
 CC neuroectodermal and pineal tumors, visual pathway and hypothalamic  
 CC glioma, Wilms' tumor or other childhood kidney tumor. The cancer is  
 CC originated from or has metastasized to the bone, brain, breast, digestive  
 CC and gastrointestinal system, endocrine system, blood, lung, respiratory  
 CC system, thorax, musculoskeletal system, or skin. The cancer is selected  
 CC from breast cancer, lung cancer, prostate cancer, ovarian cancer,  
 CC esophageal cancer, bladder cancer, leukemia, colorectal cancer and head and  
 CC testicular cancer, renal cancer, hepaticoma, neuroblastoma, lymphoma,  
 CC neck cancer. The conjugate is useful for identifying kinase substrates  
 CC capable of selectively accumulating in a target system. The conjugate is  
 CC useful for identifying conjugates capable of exhibiting selective  
 CC toxicity against a target system. The conjugate is useful for enhancing  
 CC drug efficiency. The present sequence is kinase substrate peptide useful  
 CC in the conjugate of the invention.  
 XX  
 SQ Sequence 13 AA;  
 Query Match 100.0%; Score 27; DB 9; Length 13;  
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GRGR 5  
 |||||

Db 1 GRRGR 5

RESULT 71

ADY82757

ID ADY82757 standard; peptide; 15 AA.

XX AC ADY82757;

XX DT 02-JUN-2005 (first entry)

XX DE Protein kinase C (PKC) substrate peptide - SEQ ID 30.

XX KW enzyme engineering; phosphorylation; protein kinase C; PKC; substrate.

XX OS Synthetic.

XX PN US2005064507-A1.

XX PD 24-MAR-2005.

XX PF 11-SEP-2003; 2003US-00660370.

XX PR 11-SEP-2003; 2003US-00660370.

XX PA (SHAW/) SHAW J S.

XX PI Shaw JS;

XX DR WPI; 2005-252669/26.

XX PT Test set useful for characterizing substrate specificities of kinases, comprises two peptide pools, in which every peptide in each of the peptide pools has phosphorylatable, query, anchor and degenerate amino acid positions.

XX PS Disclosure; SEQ ID NO 30; 201pp; English.

XX CC The invention comprises a test set for characterizing substrate specificities of kinases, consisting of two peptide pools, in which every peptide in each of the peptide pools contains one phosphorylatable amino acid position, one query amino acid position, at least one anchor amino acid position, and at least one degenerate amino acid position. The test set of the invention is useful to determine the spectrum of peptidyl sequences that are phosphorylated by a kinase. The present amino acid sequence represents a peptide that was used as a substrate for a protein kinase C (PKC).

XX SQ Sequence 15 AA;

PT Query Match 100.0%; Score 27; DB 9; Length 15;

PT Best Local Similarity 100.0%; Pred. No. 2.7e+02;

PT Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX QY 1 GRRGR 5

XX Db 11 GRRGR 15

RESULT 72

ADY82836

ID ADY82836 standard; peptide; 15 AA.

XX AC ADY82836;

XX DT 02-JUN-2005 (first entry)

XX DE Protein kinase C (PKC) substrate peptide - SEQ ID 109.

XX KW enzyme engineering; phosphorylation; protein kinase C; PKC; substrate.

XX OS Synthetic.

PN US2005064507-A1.

XX 24-MAR-2005.

XX PF 11-SEP-2003; 2003US-00660370.

XX PR 11-SEP-2003; 2003US-00660370.

XX PA (SHAW/) SHAW J S.

XX PI Shaw JS;

XX DR WPI; 2005-252669/26.

XX PT Test set useful for characterizing substrate specificities of kinases, comprises two peptide pools, in which every peptide in each of the peptide pools has phosphorylatable, query, anchor and degenerate amino acid positions.

XX PS Example 4; SEQ ID NO 109; 201pp; English.

XX CC The invention comprises a test set for characterizing substrate specificities of kinases, consisting of two peptide pools, in which every peptide in each of the peptide pools contains one phosphorylatable amino acid position, one query amino acid position, at least one anchor amino acid position, and at least one degenerate amino acid position. The test set of the invention is useful to determine the spectrum of peptidyl sequences that are phosphorylated by a kinase. The present amino acid sequence represents a peptide that was used as a substrate for a protein kinase C (PKC).

XX SQ Sequence 15 AA;

PT Query Match 100.0%; Score 27; DB 9; Length 15;

PT Best Local Similarity 100.0%; Pred. No. 2.7e+02;

PT Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX QY 1 GRRGR 5

XX Db 11 GRRGR 15

RESULT 73

AEB00041

ID AEB00041 standard; peptide; 15 AA.

XX AC AEB00041;

XX DT 25-AUG-2005 (first entry)

XX DE Novel dengue virus antibody-related 3C1 heavy chain CDR3 peptide SeqID39.

XX KW antibody engineering; flavivirus infection; diagnostic; pharmaceutical; virucide; antiviral; 3C1.

XX OS Pan troglodytes.

XX OS Synthetic.

XX PN WO2005056600-A2.

XX PD 23-JUN-2005.

XX PF 03-DEC-2004; 2004WO-US040674.

XX PR 08-DEC-2003; 2003US-0528161P.

XX PR 04-FEB-2004; 2004US-0541676P.

XX PR 12-MAR-2004; 2004US-0552528P.

XX PR 26-MAY-2004; 2004US-0574492P.

XX PR 01-NOV-2004; 2004US-0624261P.

XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX PI Lai C, Purcell RH;

```

XX DR WPI; 2005-458565/46.
XX
XX Novel polypeptide comprising fully human or humanized Chimpanzee
PT monoclinal antibody that binds or neutralizes dengue type 1, 2, 3, and/or
PT 4 virus, useful for treatment and diagnosis of dengue.
XX
XX Claim 4; SEQ ID NO 39; 121pp; English.
XX
XX This invention relates to novel substantially pure polypeptide comprising
CC a fully human or humanized chimpanzee monoclinal antibody that binds or
CC neutralizes dengue type 1, 2, 3, and/or 4 virus, a monoclinal antibody
CC that binds the antigen to which monoclinal antibody 5H2 (ATCC Accession
CC No. PTA-5662) binds, or a monoclinal antibody that binds the antigen to
CC which monoclinal antibody 1A5 (ATCC Accession No. PTA-6265) binds. The
CC invention may be useful for the development of compounds with a virucide
CC activity acting by viral neutralization. The invention is useful as
CC pharmaceutical preparation and a diagnostic preparation for treatment and
CC diagnosis of dengue virus disease. The present sequence is that of a 3c1
CC antibody of the invention.
XX
XX Sequence 15 AA;
SQ
Query Match 100.0%; Score 27; DB 9; Length 15;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GRGR 5
Db |||||
3 GRGR 7
RESULT 74
ABP82107
ID ABP82107 standard; peptide; 16 AA.
XX
XX ABP82107;
AC
XX
XX 04-MAR-2003 (first entry)
DT
XX
XX G protein-coupled receptor (GPCR) antigenic peptide SEQ ID NO:780.
DE
XX
XX G protein-coupled receptor; GPCR; antigenic peptide; gene therapy;
XX G protein-coupled receptor modulator; antibody; immune-related disease;
XX growth-related disease; cell regeneration-related disease; AIDS; cancer;
XX immunological-related cell proliferative disease; autoimmune disease;
XX Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;
XX osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;
XX graft versus host disease; Parkinson's disease; multiple sclerosis; pain;
XX psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;
XX mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;
XX hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;
XX ulcer.
XX
XX Homo sapiens.
OS
XX
XX WO200261087-A2.
PN
XX
XX 08-AUG-2002.
PD
XX
XX 19-DEC-2001; 2001WO-US050107.
PF
XX
XX 19-DEC-2000; 2000US-0257144P.
PR
XX
XX (LIFE-) LIFESPAN BIOSCIENCES INC.
XX
XX Burmer GC, Rough CL, Brown JP;
XX
XX WPI; 2003-046718/04.
XX
XX New isolated antigenic peptides e.g., for G protein-coupled receptors
PT (GPCR), useful for diagnosing and designing drugs for treating conditions
PT

```

```

PT in which GPCRs are involved, e.g. AIDS, Alzheimer's disease, cancer or
PT autoimmune diseases.
XX
XX Claim 1; Fig 2; 523pp; English.
XX
XX The present invention describes antigenic peptides (I) comprising: (a)
CC any one of 1601 sequences (see ABP82019 to ABP83619) of 12-24 amino
CC acids. Also described: (1) an assay for the detection of a particular G
CC protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;
CC and (2) an isolated antibody having high specificity and high affinity or
CC avidity for a particular GPCR. (I) can be used as GPCR modulators and in
CC gene therapy. The antigenic peptides for GPCRs are useful in detecting an
CC antibody against a particular GPCR, and in the production of specific
CC antibodies. The peptides and antibodies are also useful for detecting the
CC presence or absence of corresponding GPCRs. The antigenic peptides for
CC GPCRs and antibodies are useful for diagnosing and designing drugs for
CC treating immune-related diseases, growth-related diseases, cell
CC regeneration-related disease, immunological-related cell proliferative
CC diseases, or autoimmune diseases, e.g. AIDS, Alzheimer's disease,
CC atherosclerosis, bacterial, fungal, protozoan or viral infections,
CC osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute
CC inflammation, allergies, Crohn's disease, diabetes, graft versus host
CC disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,
CC anxiety, depression, schizophrenia, dementia, mental retardation, memory
CC loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension, or
CC hypotension, renal disorders, rheumatoid arthritis, trauma, ulcers, or
CC any other disorder in which GPCRs are involved. The antibodies may be
CC used in immunoassays and immunodiagnosis. ABP42523 to ABP42869 encode
CC GPCR proteins given in ABP81675 to ABP82018, which are used in the
CC exemplification of the present invention
XX
XX Sequence 16 AA;
SQ
Query Match 100.0%; Score 27; DB 6; Length 16;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GRGR 5
Db |||||
6 GRGR 10
RESULT 75
AA27504
ID AA27504 standard; peptide; 18 AA.
XX
XX AA27504;
AC
XX
XX 29-NOV-1999 (first entry)
DT
XX
XX E. coli beta'-subunit conserved region A derived peptide 39.
DE
XX
XX Bacterial life cycle; protein subunit; ribonucleic acid polymerase; RNAP;
XX enzyme function; anti-bacterial; E. coli.
XX
XX Synthetic.
OS
XX
XX Escherichia coli.
XX
XX WO9943338-A1.
PN
XX
XX 02-SEP-1999.
PD
XX
XX 26-FEB-1999; 99WO-US004351.
PF
XX
XX 26-FEB-1998; 98US-00031122.
PR
XX
XX (GLIN/) GLINSKII G V.
XX
XX Glinskii GV;
PI
XX
XX WPI; 1999-550829/46.
DR
XX
XX Identifying antibacterial drugs, by identifying compounds that block the
PT

```

PT binding of protein subunits of ribonucleic acid polymerase.  
XX  
PS  
XX Claim 27; Page 63; 106pp; English.  
CC The invention relates to methods of interfering with bacterial life cycle  
CC by bringing bacterial cells into contact with a compound that blocks the  
CC binding of at least one protein subunit of ribonucleic acid polymerase  
CC (RNAP) to a second protein subunit of RNAP. The methods can be used for  
CC obtaining compounds which inhibit subunit-subunit interactions and  
CC assembly necessary for enzyme function in bacteria. The compounds inhibit  
CC the binding of (a) at least one protein subunit of RNAP to a second  
CC protein subunit of RNAP; (b) the sigma-subunit of RNAP to the RNAP core;  
CC and blocks a nucleic acid binding to the beta-subunit or the beta'-  
CC subunit of RNAP. The compounds obtained can be used as anti-bacterial  
CC drugs. Sequences AAY27499-506 represent peptides derived from the  
CC putative nucleic acid binding sequences of the conserved region A of E.  
CC coli beta' -subunit. The antibacterial compounds that block nucleic acid  
CC binding to the beta'-subunit of the RNAP bind to the sequences indicated  
CC above  
XX  
SQ Sequence 18 AA;  
Query Match 100.0%; Score 27; DB 2; Length 18;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GRRGR 5  
DB 13 GRRGR 17  
Search completed: December 2, 2005, 10:02:15  
Job time : 169.5 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2005, 09:42:09 ; Search time 13 Seconds  
(without alignments)  
37.006 Million cell updates/sec

Title: SEQ-GRRGR  
Perfect score: 27  
Sequence: 1 grrgr 5

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 1000 summaries

Database : PIR 80:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	27	100.0	37	2	S29829
2	27	100.0	45	2	C58208
3	27	100.0	45	2	D58208
4	27	100.0	45	2	B58208
5	27	100.0	47	2	F58208
6	27	100.0	51	2	C61510
7	27	100.0	54	2	I40314
8	27	100.0	58	2	S34045
9	27	100.0	58	2	A58208
10	27	100.0	59	2	D90140
11	27	100.0	66	2	S15538
12	27	100.0	70	2	E45186
13	27	100.0	75	2	S38675
14	27	100.0	80	2	T30353
15	27	100.0	81	2	B29585
16	27	100.0	82	1	G70186
17	27	100.0	85	2	T42291
18	27	100.0	97	2	C27176
19	27	100.0	97	2	A24779
20	27	100.0	99	2	A47056
21	27	100.0	106	2	A72581
22	27	100.0	112	2	T35057
23	27	100.0	122	2	S72797
24	27	100.0	129	2	G81206
25	27	100.0	129	2	T46388
26	27	100.0	131	2	B72557
27	27	100.0	132	2	S47022
28	27	100.0	132	2	S75416
29	27	100.0	137	1	JC4877

30	27	100.0	137	1	RSBY59	ribosomal protein
31	27	100.0	137	2	S30002	ribosomal protein
32	27	100.0	137	2	S22312	ribosomal protein
33	27	100.0	137	2	D75171	ssu ribosomal prot
34	27	100.0	137	2	F71043	probable ribosomal
35	27	100.0	138	2	S46643	ribosomal protein
36	27	100.0	139	2	T38751	40S ribosomal prot
37	27	100.0	139	2	A24154	85K major surface
38	27	100.0	149	2	A30097	ribosomal protein
39	27	100.0	150	2	S11667	ribosomal protein
40	27	100.0	150	2	B30097	ribosomal protein
41	27	100.0	150	2	D84777	40S ribosomal prot
42	27	100.0	150	2	T08441	ribosomal protein
43	27	100.0	150	2	T07974	ribosomal protein
44	27	100.0	151	1	R3RT14	ribosomal protein
45	27	100.0	151	1	R4HY14	ribosomal protein
46	27	100.0	151	2	S43296	bone morphogenetic
47	27	100.0	151	2	A30815	ribosomal protein
48	27	100.0	151	2	A25220	ribosomal protein
49	27	100.0	151	2	JE0129	ribosomal protein
50	27	100.0	152	1	R3RT18	ribosomal protein
51	27	100.0	152	1	S30393	ribosomal protein
52	27	100.0	152	2	T28833	hypothetical prote
53	27	100.0	152	2	I76666	ribosomal protein
54	27	100.0	152	2	T36954	hypothetical prote
55	27	100.0	153	2	A56064	ribosomal protein
56	27	100.0	153	2	C70958	hypothetical prote
57	27	100.0	159	2	G90133	40S ribosomal prot
58	27	100.0	162	2	T42001	hypothetical prote
59	27	100.0	163	2	T28012	hypothetical prote
60	27	100.0	169	2	AF3412	hypothetical prote
61	27	100.0	170	2	AG1946	hypothetical prote
62	27	100.0	172	2	F87649	ExbB/TolR family p
63	27	100.0	182	2	H87383	hypothetical prote
64	27	100.0	183	2	D87358	hypothetical prote
65	27	100.0	189	1	TVFF85	transforming prote
66	27	100.0	189	2	S35097	transforming prote
67	27	100.0	190	2	C87660	hypothetical prote
68	27	100.0	197	2	B81974	probable transposa
69	27	100.0	197	2	A81840	probable transposa
70	27	100.0	197	2	AE2808	succinoglycan bios
71	27	100.0	198	2	G72489	hypothetical prote
72	27	100.0	200	2	B81846	insertion element
73	27	100.0	216	2	AH0120	probable membrane
74	27	100.0	216	2	T47329	hypothetical prote
75	27	100.0	217	2	G81087	IS1016C2 transposa
76	27	100.0	217	2	E81791	insertion element
77	27	100.0	217	2	B81921	probable transposa
78	27	100.0	217	2	F81964	insertion element
79	27	100.0	220	2	B96595	unknown protein, 4
80	27	100.0	222	2	A81182	IS1016C2 transposa
81	27	100.0	222	2	B81824	insertion element
82	27	100.0	224	2	C97587	succinoglycan bios
83	27	100.0	224	2	S26400	homeotic protein H
84	27	100.0	224	2	A31324	homeotic protein H
85	27	100.0	227	1	QBEC9	HXLf4 protein prec
86	27	100.0	228	2	S32563	homeotic protein H
87	27	100.0	231	2	H84500	homeotic protein H
88	27	100.0	239	2	T36765	probable ABC-type
89	27	100.0	240	2	C47751	biotin lactyl-CoA
90	27	100.0	245	1	CHUQA	complement subcomp
91	27	100.0	250	2	C87489	biotin protein lig
92	27	100.0	252	2	I46859	MHC RLA - rabbit (
93	27	100.0	256	2	F97515	birA protein (AF20
94	27	100.0	256	2	AD2734	hypothetical prote
95	27	100.0	272	2	T29446	hypothetical prote
96	27	100.0	278	2	D83474	hypothetical prote
97	27	100.0	278	2	E87704	hypothetical prote
98	27	100.0	285	2	T31503	hypothetical prote
99	27	100.0	287	2	T50647	serine/arginine-ri
100	27	100.0	287	2	AB3395	biotin-lactyl-CoA
101	27	100.0	288	2	T37029	hypothetical prote
102	27	100.0	295	2	T22833	hypothetical prote

103	27	100.0	307	2	T19582	hypothetical prote	176	27	100.0	562	2	G88408	protein nhr-6 [imp
104	27	100.0	312	2	A83112	BirA bifunctional	177	27	100.0	563	2	JQ0623	nerve growth facto
105	27	100.0	319	2	AG0457	biotin-[acetyl-CoA	178	27	100.0	567	2	G70592	probable mtrB prot
106	27	100.0	320	2	AG0934	bifunctional prote	179	27	100.0	577	2	S33763	hormone receptor N
107	27	100.0	320	2	E82337	BirA bifunctional	180	27	100.0	589	2	S71934	RNA/sDNA-binding
108	27	100.0	321	1	BV5CBF	biotin-[acetyl-CoA	181	27	100.0	589	2	F95292	probable ABC trans
109	27	100.0	321	2	A86089	hypothetical prote	182	27	100.0	598	2	A57040	T-cell nuclear rec
110	27	100.0	321	2	D91241	hypothetical prote	183	27	100.0	598	2	A37251	probable nuclear h
111	27	100.0	322	2	AF3047	conserved hypothet	184	27	100.0	598	2	A46225	nuclear orphan rec
112	27	100.0	323	2	AC2481	hypothetical prote	185	27	100.0	598	2	I84692	nuclear orphan rec
113	27	100.0	323	2	T24224	hypothetical prote	186	27	100.0	601	1	Q8MSN1	probable hormone r
114	27	100.0	323	2	D86847	biotin-[acetyl-CoA	187	27	100.0	602	2	E70773	probable rho prote
115	27	100.0	327	2	E98238	hypothetical prote	188	27	100.0	610	2	T09988	probable transcrip
116	27	100.0	335	1	H71071	hypothetical prote	189	27	100.0	614	2	T33149	hypothetical prote
117	27	100.0	337	2	H70125	PA54 protein - yea	190	27	100.0	619	1	T43348	nuclear steroid ho
118	27	100.0	340	1	WMBEL1	latency-related pr	191	27	100.0	619	2	B97475	hypothetical prote
119	27	100.0	340	2	A24026	erythromycin resis	192	27	100.0	619	2	AF2693	conserved hypothet
120	27	100.0	351	1	JSBYP1	centromere-binding	193	27	100.0	621	2	A32838	DNA-directed RNA p
121	27	100.0	358	2	A75561	conserved hypothet	194	27	100.0	624	2	A55576	collagen alpha 2(X
122	27	100.0	359	2	T01275	hypothetical prote	195	27	100.0	625	2	B42361	DNA-directed RNA p
123	27	100.0	359	2	A98148	fructose transport	196	27	100.0	625	2	AE2005	RNA polymerase gam
124	27	100.0	359	2	AC3140	transcription regu	197	27	100.0	625	2	S71930	neuron-derived rec
125	27	100.0	360	2	S58205	DHR38 protein - si	198	27	100.0	626	2	S75807	DNA-directed RNA p
126	27	100.0	370	2	S29186	vasopressin V2 rec	199	27	100.0	628	2	UC2493	neuron derived orp
127	27	100.0	370	2	S34042	vasopressin recept	200	27	100.0	629	2	S73172	DNA-directed RNA p
128	27	100.0	370	2	S68777	vasopressin recept	201	27	100.0	636	2	T06917	DNA-directed RNA p
129	27	100.0	371	2	S44045	vasopressin recept	202	27	100.0	641	1	Q0BE31	nuclear antigen BB
130	27	100.0	371	2	T48760	hypothetical prote	203	27	100.0	646	2	T35002	probable respirato
131	27	100.0	380	2	E84748	probable protein p	204	27	100.0	653	2	T49842	related to Na+/H+-
132	27	100.0	382	2	E95391	probable membrane	205	27	100.0	657	2	A54278	DNA-binding protei
133	27	100.0	383	2	G75370	hypothetical prote	206	27	100.0	678	2	A71287	probable cytoplasm
134	27	100.0	386	2	T19504	hypothetical prote	207	27	100.0	682	2	B84415	translation elonga
135	27	100.0	388	2	E96513	unknown protein, 3	208	27	100.0	694	2	S09794	hypothetical prote
136	27	100.0	402	2	T35137	hypothetical prote	209	27	100.0	728	1	S07558	translation elonga
137	27	100.0	403	2	S70676	glycosyltransferas	210	27	100.0	739	1	JQ1893	80-7K alpha trans-
138	27	100.0	414	2	T43851	translation initia	211	27	100.0	741	2	I51657	suppressor of yea
139	27	100.0	414	2	C86301	arginine/serine-ri	212	27	100.0	770	2	B56695	transducin-like en
140	27	100.0	417	2	E30341	alpha-1-adrenergic	213	27	100.0	802	2	B84560	hypothetical prote
141	27	100.0	421	2	D72677	hypothetical prote	214	27	100.0	810	2	T44430	protein pv100 [imp
142	27	100.0	424	2	S74788	hypothetical prote	215	27	100.0	813	2	G96494	protein F7F22.4 [i
143	27	100.0	427	2	T20800	hypothetical prote	216	27	100.0	822	2	T51049	related to nucleol
144	27	100.0	430	2	S66671	neuron-derived rec	217	27	100.0	859	2	S70584	nitrate reductase
145	27	100.0	436	2	JQ1459	Bt1 protein precu	218	27	100.0	867	1	JQ1525	nitrate reductase
146	27	100.0	437	2	C39135	hypothetical prote	219	27	100.0	868	1	JC4283	nitrate reductase
147	27	100.0	441	2	T02930	oxidoreductase Atu	220	27	100.0	873	1	JH0182	nitrate reductase
148	27	100.0	445	2	T03578	probable glucose-i	221	27	100.0	881	2	T26498	hypothetical prote
149	27	100.0	446	2	F98351	probable oxidoredu	222	27	100.0	883	2	C70879	probable ftsK - My
150	27	100.0	453	2	C91170	probable surfactin	223	27	100.0	893	2	S46442	nitrate reductase
151	27	100.0	453	2	C86016	probable enzyme Z4	224	27	100.0	898	2	H87481	ribonuclease, Rne/
152	27	100.0	467	1	VMUT7R	variant surface gl	225	27	100.0	900	2	C96842	hypothetical prote
153	27	100.0	474	2	D75285	BirA bifunctional	226	27	100.0	922	2	T38371	probable helicase
154	27	100.0	497	2	S74404	o-succinylbenzoate	227	27	100.0	924	2	E87092	probable ribonucle
155	27	100.0	498	1	W2WL8	E2 protein - human	228	27	100.0	940	2	A87502	hypothetical prote
156	27	100.0	499	2	S09880	hypothetical prote	229	27	100.0	946	2	T01460	hypothetical prote
157	27	100.0	501	2	E70596	hypothetical prote	230	27	100.0	953	2	B70681	probable rne prote
158	27	100.0	505	2	G86391	hypothetical prote	231	27	100.0	962	2	T00262	hypothetical prote
159	27	100.0	511	1	A48560	UL54 protein - hum	232	27	100.0	983	2	H72510	probable ribonucle
160	27	100.0	512	1	WMBEXA	UL54 protein - hum	233	27	100.0	994	2	A70776	probable glnE - My
161	27	100.0	512	1	WMBEY4	UL54 protein - hum	234	27	100.0	1000	2	T39423	hypothetical prote
162	27	100.0	513	2	A96265	hypothetical prote	235	27	100.0	1004	2	H87112	glutamate-ammonia-
163	27	100.0	513	2	AH3019	sigma 54 dependent	236	27	100.0	1011	2	T07712	probable ABC-type
164	27	100.0	515	2	A40491	alpha-1B-adrenergic	237	27	100.0	1027	2	I38604	hypothetical protei
165	27	100.0	515	2	JC1525	alpha-1B-adrenergic	238	27	100.0	1051	2	T51904	hypothetical prote
166	27	100.0	517	2	A45121	alpha-1B adrenergi	239	27	100.0	1053	2	T09645	DNA-directed RNA p
167	27	100.0	521	2	B86332	hypothetical prote	240	27	100.0	1053	2	T09641	DNA-directed RNA p
168	27	100.0	531	2	T08760	hypothetical prote	241	27	100.0	1055	2	T10432	DNA-directed RNA p
169	27	100.0	536	2	T42606	probable transcrip	242	27	100.0	1057	2	T10908	DNA-directed RNA p
170	27	100.0	538	2	H86335	T20H2.2 protein -	243	27	100.0	1067	2	T18196	pol protein - silk
171	27	100.0	539	2	A87382	hypothetical prote	244	27	100.0	1145	2	T51546	cellulose synthase
172	27	100.0	543	2	S46713	ATP-dependent RNA	245	27	100.0	1176	2	A49848	nitrite reductase
173	27	100.0	545	2	T01288	protein kinase F27	246	27	100.0	1181	2	D86157	hypothetical prote
174	27	100.0	546	2	T02029	DNA-binding protei	247	27	100.0	1182	2	E97286	DNA-dependent RNA
175	27	100.0	551	2	S58238	DHR38 protein - fr	248	27	100.0	1188	2	G72734	hypothetical prote

249	27	100.0	1189	2	I39711	celE protein - Agr	322	24	88.9	172	2	T49357	hypothetical prote
250	27	100.0	1193	2	A86193	hypothetical prote	323	24	88.9	177	2	D95278	conserved hypothet
251	27	100.0	1198	2	G86849	hypothetical prote	324	24	88.9	179	2	G81086	insertion element
252	27	100.0	1199	2	G69698	RNA polymerase (be	325	24	88.9	182	2	T01529	hypothetical prote
253	27	100.0	1201	2	AD1107	RNA polymerase (be	326	24	88.9	182	2	T31182	hypothetical prote
254	27	100.0	1201	2	AG1468	RNA polymerase (be	327	24	88.9	188	2	F97428	hypothetical prote
255	27	100.0	1206	2	T44376	DNA-directed RNA p	328	24	88.9	194	2	G64422	ribosomal protein
256	27	100.0	1207	2	A89822	RNA polymerase bet	329	24	88.9	201	2	T16878	hypothetical prote
257	27	100.0	1225	2	B95229	DNA-directed RNA p	330	24	88.9	218	2	A34445	25K calcium-bindin
258	27	100.0	1225	2	F98093	DNA-directed RNA p	331	24	88.9	219	2	F71034	hypothetical prote
259	27	100.0	1245	2	H87340	hypothetical prote	332	24	88.9	220	2	T50624	hypothetical prote
260	27	100.0	1274	2	D84485	probable retroelem	333	24	88.9	221	2	D75179	hypothetical prote
261	27	100.0	1284	2	T40879	Probable helicase	334	24	88.9	227	2	T49711	hypothetical prote
262	27	100.0	1316	2	D87145	[beta]' subunit of	335	24	88.9	230	1	B8AG6	virB8 protein - Ag
263	27	100.0	1316	2	S31146	DNA-directed RNA p	336	24	88.9	230	2	A13470	two component resp
264	27	100.0	1316	2	G70535	probable rpoC prot	337	24	88.9	231	2	S50382	hypothetical prote
265	27	100.0	1356	2	T16718	hypothetical prote	338	24	88.9	233	2	B97478	popp protein [limpo
266	27	100.0	1371	2	S77521	sensory transducti	339	24	88.9	240	2	G83271	probable phosphodi
267	27	100.0	1372	2	F97722	hypothetical prote	340	24	88.9	244	1	TVWVFB	transforming prote
268	27	100.0	1372	2	B71724	dna-directed RNA p	341	24	88.9	246	2	E83249	probable short-cha
269	27	100.0	1380	2	S64721	protoporphylin IX	342	24	88.9	257	1	B8AG55	virB8 protein prec
270	27	100.0	1389	2	F58157	peritaxin - rat	343	24	88.9	257	2	E84356	dolichol-P-glucose
271	27	100.0	1396	2	F87311	DNA-directed RNA p	344	24	88.9	261	2	H69122	biotin acetyl-CoA
272	27	100.0	1399	2	G83112	DNA-directed RNA p	345	24	88.9	268	2	F72748	probable fkbp-type
273	27	100.0	1400	2	AH3345	DNA-directed RNA p	346	24	88.9	270	2	F86170	GBF4 [imported] -
274	27	100.0	1401	2	G82336	DNA-directed RNA p	347	24	88.9	273	2	E96015	probable acetate o
275	27	100.0	1402	2	A12816	DNA-directed RNA p	348	24	88.9	274	2	B95112	formamidopyrimidin
276	27	100.0	1406	2	AB0456	DNA-directed RNA p	349	24	88.9	274	2	H84043	formamidopyrimidin
277	27	100.0	1407	1	RNECC	DNA-directed RNA p	350	24	88.9	274	2	H97980	DNA-formamidopyrim
278	27	100.0	1407	2	H84933	DNA-directed RNA p	351	24	88.9	276	2	A49853	ammonia monooxygen
279	27	100.0	1407	2	AD0933	DNA-directed RNA p	352	24	88.9	277	2	AD1961	hypothetical prote
280	27	100.0	1407	2	G91242	RNA polymerase bet	353	24	88.9	277	2	T21135	hypothetical prote
281	27	100.0	1407	2	D86090	RNA polymerase, be	354	24	88.9	278	2	B69663	formamidopyrimidin
282	27	100.0	1415	1	G64073	DNA-directed RNA p	355	24	88.9	282	2	AG2971	beta-lactamase [im
283	27	100.0	1416	2	B97595	hypothetical prote	356	24	88.9	283	2	E69491	conserved hypothet
284	27	100.0	1430	2	D82533	RNA polymerase bet	357	24	88.9	288	2	C98311	probable beta-lact
285	27	100.0	1520	2	T00273	hypothetical prote	358	24	88.9	291	2	S51608	RYB-a protein - ra
286	27	100.0	1528	2	T06694	hypothetical prote	359	24	88.9	291	2	S48055	RYB-a protein - ra
287	27	100.0	1546	1	CGH02E	collagen alpha 2(X	360	24	88.9	295	2	F95997	probable transcrip
288	27	100.0	1547	2	T13847	sno protein - frui	361	24	88.9	297	2	T33914	hypothetical prote
289	27	100.0	1557	2	T13160	protein CNK - frui	362	24	88.9	297	2	T18638	hypothetical prote
290	27	100.0	1634	2	T26517	hypothetical prote	363	24	88.9	301	2	C84906	probable rac GTPas
291	27	100.0	1838	1	CGH01V	collagen alpha 1(V	364	24	88.9	302	2	A33964	regulatory protein
292	27	100.0	1843	2	S18803	collagen alpha 1(V	365	24	88.9	307	2	B95417	probable ABC trans
293	27	100.0	2925	2	T00133	DNA-directed RNA p	366	24	88.9	312	2	C70901	probable fnt prote
294	24	88.9	29	2	JH0699	omega-conotoxin MV	367	24	88.9	318	2	H86977	probable methionyl
295	24	88.9	48	2	A57125	polypheolic adhes	368	24	88.9	322	1	TVFVF4	transforming prote
296	24	88.9	79	2	A95090	KH domain protein	369	24	88.9	324	1	JC4985	UTP-glucose-1-phos
297	24	88.9	79	2	C97957	conserved hypothet	370	24	88.9	330	2	JC5717	G protein-coupled
298	24	88.9	91	2	A59493	protamine P2 - Sty	371	24	88.9	330	2	T34972	probable histone d
299	24	88.9	101	2	H90223	ribosomal protein	372	24	88.9	334	2	B75095	LPS glycosyltransf
300	24	88.9	105	2	A27471	homeotic protein R	373	24	88.9	346	2	G69325	hypothetical prote
301	24	88.9	108	2	S05258	probable protein H	374	24	88.9	349	2	T49843	hypothetical prote
302	24	88.9	124	2	T37118	probable transposa	375	24	88.9	349	2	F85294	splicing factor At
303	24	88.9	130	2	D69146	ribosomal protein	376	24	88.9	350	2	T05797	splicing factor SR
304	24	88.9	133	2	C69535	SSU ribosomal prot	377	24	88.9	356	2	F89977	hypothetical prote
305	24	88.9	136	2	S51244	hypothetical prote	378	24	88.9	357	2	S35318	MET22 protein - ye
306	24	88.9	138	2	E48188	phospholipase A2 (	379	24	88.9	361	2	T45778	hypothetical prote
307	24	88.9	138	2	D48188	phospholipase A2 (	380	24	88.9	364	2	AB1918	protein phosphatas
308	24	88.9	142	2	D32880	hypothetical prote	381	24	88.9	367	1	TVCHFS	transforming prote
309	24	88.9	145	2	A99112	40S ribosomal prot	382	24	88.9	370	2	A60089	transforming prote
310	24	88.9	145	2	G75031	hypothetical prote	383	24	88.9	372	2	T01600	hypothetical prote
311	24	88.9	146	2	E69535	SSU ribosomal prot	384	24	88.9	374	2	E90029	hypothetical prote
312	24	88.9	147	2	AG2646	conserved hypothet	385	24	88.9	375	2	AH3110	oxoreductase [limpo
313	24	88.9	148	2	B71137	hypothetical prote	386	24	88.9	375	2	D98176	hypothetical prote
314	24	88.9	150	2	S28049	histone H2b - maiz	387	24	88.9	377	2	G97591	preproteins transio
315	24	88.9	151	2	D82030	probable IS1016 tr	388	24	88.9	377	2	AH3328	NAD(FAD)-utilizing
316	24	88.9	151	2	D81818	insertion element	389	24	88.9	379	2	H69512	hypothetical prote
317	24	88.9	152	1	S37496	ribosomal protein	390	24	88.9	380	1	TVHUF1	transforming prote
318	24	88.9	152	2	T39575	ribosomal protein	391	24	88.9	380	1	TVMSF	transforming prote
319	24	88.9	153	2	S47943	translation initia	392	24	88.9	380	1	TVRTFS	transforming prote
320	24	88.9	154	2	T27228	ribosomal protein	393	24	88.9	381	1	TVWVJ	transforming prote
321	24	88.9	165	2	A59492	protamine P1 - Sty	394	24	88.9	383	2	AB2948	conserved hypothet

395	24	88.9	303	2	H98334	hypothetical prote	468	24	88.9	739	2	T52289	probable transketo
396	24	88.9	387	2	H87541	conserved hypotet	469	24	88.9	750	2	T48804	hypothetical prote
397	24	88.9	389	2	G83413	probable MFS trans	470	24	88.9	763	2	T08929	hypothetical prote
398	24	88.9	390	1	E64099	GTP-binding protei	471	24	88.9	765	2	T49592	neurofilament trip
399	24	88.9	393	2	E87432	conserved hypotet	472	24	88.9	767	2	S64571	probable membrane
400	24	88.9	394	2	AF2650	conserved hypotet	473	24	88.9	771	2	A80823	penicillin-binding
401	24	88.9	401	2	A71199	hypothetical prote	474	24	88.9	774	2	AC0335	probable enoyl-CoA
402	24	88.9	405	2	A75105	hypothetical prote	475	24	88.9	780	2	T96840	hypothetical prote
403	24	88.9	415	2	H83067	probable c-type cy	476	24	88.9	800	2	B87675	tonB-dependent rec
404	24	88.9	420	2	E97432	hypothetical prote	477	24	88.9	803	2	AD1282	leucyl-tRNA synthe
405	24	88.9	422	2	E83869	fatty acid alpha h	478	24	88.9	803	2	AH1653	leucyl-tRNA synthe
406	24	88.9	424	2	F69476	acetylornithine am	479	24	88.9	814	2	JC7389	thyroid stimulatn
407	24	88.9	424	2	T39524	hypothetical prote	480	24	88.9	825	2	JC4163	DNA-binding protei
408	24	88.9	434	2	A82145	NADH dehydrogenase	481	24	88.9	837	2	T49562	neurofilament-H re
409	24	88.9	435	2	D71082	hypothetical prote	482	24	88.9	845	2	D96799	hypothetical prote
410	24	88.9	437	2	B75067	hypothetical prote	483	24	88.9	847	2	C96703	hypothetical prote
411	24	88.9	445	2	A75376	probable oligosacc	484	24	88.9	864	2	H85335	hypothetical prote
412	24	88.9	445	2	T50802	serine/threonine p	485	24	88.9	864	2	T04518	large helicase-rel
413	24	88.9	446	2	AE2813	Preprotein translo	486	24	88.9	867	2	D75091	hypothetical prote
414	24	88.9	449	2	S49173	variable surface g	487	24	88.9	887	2	B96491	hypothetical prote
415	24	88.9	458	2	G83690	hypothetical prote	488	24	88.9	896	2	D95418	probable sensor pr
416	24	88.9	459	2	G86212	hypothetical prote	489	24	88.9	898	2	A82298	translation initia
417	24	88.9	460	2	A12218	hypothetical prote	490	24	88.9	940	2	T01854	hypothetical prote
418	24	88.9	467	2	AC3324	UDP-N-acetylmuram	491	24	88.9	951	2	T45726	collagen alpha 3(V
419	24	88.9	468	2	AE2422	hypothetical prote	492	24	88.9	959	2	S32605	collagen alpha 3(V
420	24	88.9	469	2	B84644	probable protein k	493	24	88.9	973	2	JQ0971	transposase tnpA -
421	24	88.9	471	2	S15035	acetylcholinestera	494	24	88.9	995	2	T51051	hypothetical prote
422	24	88.9	472	2	T51456	pectate lyase-like	495	24	88.9	996	2	B84481	hypothetical prote
423	24	88.9	474	2	S66480	carbon catabolite	496	24	88.9	1016	2	D86308	translation initia
424	24	88.9	476	2	B96547	probable protein k	497	24	88.9	1024	2	S18251	collagen alpha 1(X
425	24	88.9	482	2	T15217	hypothetical prote	498	24	88.9	1026	2	T26282	hypothetical prote
426	24	88.9	486	2	T35375	probable membrane	499	24	88.9	1028	2	A96719	hypothetical prote
427	24	88.9	504	2	D17248	regulatory protein	500	24	88.9	1036	2	D84741	probable cellulose
428	24	88.9	505	2	S95970	probable sugar upt	501	24	88.9	1051	2	D87349	AcxB/AcrD/AcrF fam
429	24	88.9	509	2	AD0415	dihydrolipoamide S	502	24	88.9	1076	2	A69409	carbamoyl-phosphat
430	24	88.9	514	2	T35371	probable Na+/H+ an	503	24	88.9	1088	2	T14917	homeotic protein p
431	24	88.9	520	2	T06625	hypothetical prote	504	24	88.9	1111	2	T05646	hypothetical prote
432	24	88.9	522	2	AB3426	ribose transport A	505	24	88.9	1113	1	A47106	myosin heavy chain
433	24	88.9	528	2	B84743	hypothetical prote	506	24	88.9	1121	2	T06065	hypothetical prote
434	24	88.9	534	2	S55635	DNA helicase-prima	507	24	88.9	1142	2	JX0369	collagen alpha 1(X
435	24	88.9	538	2	S57459	hook-containing pr	508	24	88.9	1200	2	T15921	hypothetical prote
436	24	88.9	547	2	S15028	chromatin-binding	509	24	88.9	1299	2	T43251	furin (EC 3.4.21.7
437	24	88.9	562	2	S69055	TTAGGG repeat-bind	510	24	88.9	1352	2	G71051	probable ATP-depen
438	24	88.9	566	1	T46549	isobutyryl-CoA mut	511	24	88.9	1384	2	T02748	hypothetical prote
439	24	88.9	566	2	AC0559	probable solute-bi	512	24	88.9	1391	2	F81233	DNA-directed RNA p
440	24	88.9	567	2	I64111	dihydrolipoamide S	513	24	88.9	1391	2	C82007	DNA-directed RNA p
441	24	88.9	569	2	T19483	hypothetical prote	514	24	88.9	1484	2	T29275	hypothetical prote
442	24	88.9	572	2	T34658	hypothetical prote	515	24	88.9	1560	2	T00080	hypothetical prote
443	24	88.9	581	2	T29830	hypothetical prote	516	24	88.9	1596	2	T31338	cellulose synthase
444	24	88.9	591	2	A69159	protoporphyrin IX	517	24	88.9	1670	1	CGHU3B	collagen alpha 3(I
445	24	88.9	606	2	F69493	conserved hypotet	518	24	88.9	1744	2	S40991	collagen alpha 1(I
446	24	88.9	623	2	A45137	collagen alpha 4(I	519	24	88.9	1806	1	CGHU1E	collagen alpha 1(X
447	24	88.9	627	2	T02846	dynein light chain	520	24	88.9	1930	2	F86200	protein F12K11.17
448	24	88.9	628	2	AC87096	probable GTP-bind	521	24	88.9	2025	2	T03884	hypothetical prote
449	24	88.9	629	2	AH0521	dihydrolipoamide a	522	24	88.9	2165	2	T21371	hypothetical prote
450	24	88.9	630	1	XXECDP	dihydrolipoamide S	523	24	88.9	2492	1	A44213	nonstructural poly
451	24	88.9	630	2	G90643	hypothetical prote	524	24	88.9	2492	1	C44213	nonstructural poly
452	24	88.9	630	2	G85494	hypothetical prote	525	24	88.9	2492	1	MNWVTD	nonstructural poly
453	24	88.9	633	2	T24898	hypothetical prote	526	24	88.9	2514	1	MNWVN2	polyketide synthas
454	24	88.9	633	2	S36723	FUN36 protein - ye	527	24	88.9	2543	2	F69679	probable acetyl-Co
455	24	88.9	643	2	D70697	probable membranep	528	24	88.9	2655	2	D96595	collagen alpha 3(V
456	24	88.9	644	2	T47835	hypothetical prote	529	24	88.9	3176	2	CGHU3A	collagen alpha 3(V
457	24	88.9	646	2	T09906	hypothetical prote	530	24	88.9	3396	2	T22613	collagen alpha 3(V
458	24	88.9	646	2	T02398	hypothetical prote	531	23	85.2	55	2	A37238	autoimmune epitope
459	24	88.9	656	2	B47096	hlyB homolog - Str	532	23	85.2	57	2	T03643	hypothetical prote
460	24	88.9	670	2	S53414	probable membrane	533	23	85.2	71	2	T42025	hypothetical prote
461	24	88.9	676	2	T01084	hypothetical prote	534	23	85.2	79	2	H66820	conserved hypotet
462	24	88.9	687	2	T49960	FtsH-like protein	535	23	85.2	82	2	F75326	probable excisiona
463	24	88.9	702	2	S41685	NGG1 protein - yea	536	23	85.2	93	2	AB0740	ribosomal protein
464	24	88.9	703	2	D71420	hypothetical prote	537	23	85.2	97	2	H69440	Hypothetical prote
465	24	88.9	712	2	S68306	pol polyprotein, t	538	23	85.2	106	2	AG2867	NADH2 dehydrogenas
466	24	88.9	717	2	H85171	DEF (CtAl) protein	539	23	85.2	120	2	S27227	probable small rib
467	24	88.9	728	2	A54603	transcription fact	540	23	85.2	121	2	T40190	

541	23	85.2	122	2	F84180	hypothetical prote	614	23	85.2	403	2	A91097	probable deacetyla
542	23	85.2	124	2	S54782	PolII transcriptio	615	23	85.2	403	2	H65070	hypothetical prote
543	23	85.2	126	2	T51113	precortin-6y C5,15	616	23	85.2	403	2	E85942	probable deacetyla
544	23	85.2	126	2	AD3508	transcription regu	617	23	85.2	405	2	I39062	dematin 52K chain
545	23	85.2	136	2	T48713	hypothetical prote	618	23	85.2	412	2	S62538	hypothetical coile
546	23	85.2	138	2	PC1190	genome polyprotein	619	23	85.2	413	2	B83253	probable glycosyl
547	23	85.2	138	2	PC1189	genome polyprotein	620	23	85.2	437	2	A47104	chloride channel 6
548	23	85.2	138	2	PC1191	genome polyprotein	621	23	85.2	440	2	T50912	hypothetical prote
549	23	85.2	145	2	C84498	nifZ protein-like trans	622	23	85.2	441	2	C70702	probable membrane
550	23	85.2	148	2	S02509	nifZ protein - K1e	623	23	85.2	445	1	G70647	NADH2 dehydrogenas
551	23	85.2	154	2	B83304	hypothetical prote	624	23	85.2	449	2	T34619	NADH2 dehydrogenas
552	23	85.2	174	2	E78552	ribosomal protein	625	23	85.2	450	2	G01158	tyrosine kinase ac
553	23	85.2	187	2	E69170	hypothetical prote	626	23	85.2	452	2	T30082	hypothetical prote
554	23	85.2	190	2	S35388	heat shock protein	627	23	85.2	453	2	AB1867	GTP binding protei
555	23	85.2	198	2	S76554	hypothetical prote	628	23	85.2	461	2	S52408	SPVR1 protein - s
556	23	85.2	204	2	C40899	hypothetical prote	629	23	85.2	464	2	S10171	3-isopropylmalate
557	23	85.2	206	2	E83113	30S ribosomal prot	630	23	85.2	466	2	AD0516	3-isopropylmalate
558	23	85.2	211	2	B37499	glial cell line-de	631	23	85.2	466	2	D90638	hypothetical prote
559	23	85.2	211	2	I49686	glial cell line-de	632	23	85.2	466	2	D85489	hypothetical prote
560	23	85.2	211	2	A37499	glial cell line-de	633	23	85.2	466	2	H64728	3-isopropylmalate
561	23	85.2	212	2	B70301	ribosomal protein	634	23	85.2	467	2	H82070	3-isopropylmalate
562	23	85.2	212	2	C75636	probable N-acetyl m	635	23	85.2	469	2	F64163	3-isopropylmalate
563	23	85.2	216	2	F71366	probable heat-shoc	636	23	85.2	476	2	AI0065	3-isopropylmalate
564	23	85.2	228	2	A82168	conserved hypotet	637	23	85.2	477	2	H75267	hypothetical prote
565	23	85.2	237	2	T27850	hypothetical prote	638	23	85.2	479	2	C87273	3-isopropylmalate
566	23	85.2	240	2	C43649	hypothetical prote	639	23	85.2	487	2	AC7471	testis zinc finger
567	23	85.2	246	2	AE2157	hypothetical prote	640	23	85.2	492	2	A97802	n utilization subs
568	23	85.2	248	2	G69111	conserved hypotet	641	23	85.2	503	2	A97802	n utilization subs
569	23	85.2	250	2	B35026	filaggrin B - mous	642	23	85.2	503	2	H71659	transcription term
570	23	85.2	252	2	B64875	probable dehydroge	643	23	85.2	506	1	W2ML47	E2 protein - human
571	23	85.2	252	2	D85760	probable oxidoredu	644	23	85.2	508	1	DTEBPH	purH bifunctional
572	23	85.2	252	2	C90859	probable oxidoredu	645	23	85.2	529	1	DTECPH	purH bifunctional
573	23	85.2	252	2	C72781	probable uroporphyr	646	23	85.2	529	2	AB0931	hypothetical prote
574	23	85.2	253	2	AC0654	hypothetical oxido	647	23	85.2	529	2	A98245	hypothetical prote
575	23	85.2	253	2	AI0269	probable short cha	648	23	85.2	529	2	G86092	hypothetical prote
576	23	85.2	255	2	A35026	filaggrin A - mous	649	23	85.2	529	2	AH0453	bifunctional purin
577	23	85.2	259	2	T00935	hypothetical prote	650	23	85.2	530	2	C82344	phosphoribosylamin
578	23	85.2	265	1	H64151	hydroxyethylthiaz	651	23	85.2	536	2	T03034	cytochrome p450 -
579	23	85.2	266	2	G64022	hypothetical prote	652	23	85.2	536	2	A40676	sphere organelle p
580	23	85.2	267	2	A72777	hypothetical prote	653	23	85.2	537	2	AI3497	N utilization subs
581	23	85.2	269	2	A83086	conserved hypotet	654	23	85.2	538	2	AD2587	N-utilization subs
582	23	85.2	278	1	Q0BE17	dUTP diphosphatase	655	23	85.2	538	2	D97369	transcription term
583	23	85.2	280	2	T09576	Lbx1 transcription	656	23	85.2	548	2	D87254	N utilization subs
584	23	85.2	281	2	JE0418	aminoglycoside-N-a	657	23	85.2	553	1	D46329	cell fusion glycop
585	23	85.2	284	2	A07119	2-dehydro-3-deoxyp	658	23	85.2	553	1	E46329	cell fusion glycop
586	23	85.2	285	2	A97653	probable rRNA meth	659	23	85.2	566	1	HMIUSA	hemagglutinin prec
587	23	85.2	288	2	H90361	hypothetical prote	660	23	85.2	576	2	C87261	hypothetical prote
588	23	85.2	301	2	C82791	conserved hypotet	661	23	85.2	592	2	B81009	BirA protein/Bvg a
589	23	85.2	302	2	A99889	regulatory protein	662	23	85.2	592	2	H82031	probable biotin-la
590	23	85.2	312	2	I48303	collagen alpha 4(I	663	23	85.2	592	2	G70786	probable ephd prot
591	23	85.2	313	2	H82335	conserved hypotet	664	23	85.2	599	2	T25835	hypothetical prote
592	23	85.2	316	2	T19291	hypothetical prote	665	23	85.2	643	1	F0LTLK	gag polyprotein -
593	23	85.2	316	2	T19288	hypothetical prote	666	23	85.2	650	2	A54976	peroxisomal matrix
594	23	85.2	317	2	B82637	bifunctional bioti	667	23	85.2	652	2	G82962	hypothetical prote
595	23	85.2	320	2	A83519	hypothetical prote	668	23	85.2	691	2	F91251	probable tape meas
596	23	85.2	328	2	T46256	brevican - human (	669	23	85.2	694	2	AG2353	translation elonga
597	23	85.2	331	2	AC3003	site-specific reco	670	23	85.2	721	2	A39707	erythrocyte membra
598	23	85.2	331	2	F98280	integrinase/recombin	671	23	85.2	728	2	E69486	translation elonga
599	23	85.2	335	2	T05722	licheninase (EC 3.	672	23	85.2	738	2	E87627	hypothetical prote
600	23	85.2	345	2	H70722	probable transfera	673	23	85.2	759	2	D84301	cell division cycl
601	23	85.2	356	2	A33821	chromosomal protei	674	23	85.2	760	2	F75530	ribonuclease - Del
602	23	85.2	360	2	T51344	RNA helicase RH18	675	23	85.2	771	2	A87532	glucoamylase (impo
603	23	85.2	372	2	B72472	probable electron	676	23	85.2	771	2	B70564	hypothetical prote
604	23	85.2	374	2	D70573	hypothetical prote	677	23	85.2	780	2	AE0061	organic solvent to
605	23	85.2	375	2	A75264	hypothetical prote	678	23	85.2	782	1	G64157	probable organic s
606	23	85.2	378	2	D95064	N utilization subs	679	23	85.2	784	1	F64726	organic solvent to
607	23	85.2	378	2	F97931	transcription term	680	23	85.2	784	2	AI0513	organic solvent to
608	23	85.2	384	2	T29422	hypothetical prote	681	23	85.2	784	2	C85487	organic solvent to
609	23	85.2	386	2	C97644	hypothetical prote	682	23	85.2	784	2	C90636	organic solvent to
610	23	85.2	389	1	VH1H2E	nucleocapsid prote	683	23	85.2	790	1	TVHUTT	nerve growth facto
611	23	85.2	389	2	S08031	nucleocapsid prote	684	23	85.2	836	2	B96716	probable serine/th
612	23	85.2	392	2	D82206	GGDEF family prote	685	23	85.2	856	2	T31234	trac protein homol
613	23	85.2	393	2	A84600	probable fructose	686	23	85.2	875	2	T17382	vrlC protein - Dlc

687	23	85.2	928	1	VCBEBC	glycoprotein g1 pr	760	22	81.5	98	1	KEGLBS	keratin, feather -
688	23	85.2	932	1	VCBEBC	glycoprotein g1 pr	761	22	81.5	98	2	D72601	hypothetical prote
689	23	85.2	932	2	B70929	hypothetical prote	762	22	81.5	98	2	JC5065	calcium-binding pr
690	23	85.2	933	1	S56050	replication licens	763	22	81.5	98	2	D82764	hypothetical prote
691	23	85.2	1016	2	G86295	hypothetical prote	764	22	81.5	99	2	B38623	collagen alpha 2(I
692	23	85.2	1027	2	I38759	zinc finger/leucin	765	22	81.5	102	2	S72839	probable transcrip
693	23	85.2	1035	2	S18512	cell division cont	766	22	81.5	102	2	E70737	probable whiB3 pro
694	23	85.2	1051	2	A57633	collagen alpha 2 c	767	22	81.5	102	2	B20099	hypothetical prote
695	23	85.2	1106	2	JQ0405	hypothetical 119.5	768	22	81.5	103	2	G86911	hypothetical prote
696	23	85.2	1122	2	T14180	exit protein - Myc	769	22	81.5	103	2	A72713	hypothetical prote
697	23	85.2	1146	2	A38587	collagen, cornea-s	770	22	81.5	104	2	T24609	hypothetical prote
698	23	85.2	1171	2	F83110	exodeoxyribonuclea	771	22	81.5	104	2	C85679	probable holin pro
699	23	85.2	1191	2	S76414	beta transducin-li	772	22	81.5	104	2	G85819	hypothetical prote
700	23	85.2	1271	2	T49009	protein kinase lik	773	22	81.5	104	2	C87685	hypothetical prote
701	23	85.2	1295	2	T30528	reverse transcript	774	22	81.5	104	2	AC2099	hypothetical prote
702	23	85.2	1309	2	T00078	probable RNA-direc	775	22	81.5	105	2	G95371	hypothetical prote
703	23	85.2	1312	1	JC4171	peptidyl-dipectida	776	22	81.5	105	2	T16208	hypothetical prote
704	23	85.2	1313	1	AC2038	peptidyl-dipectida	777	22	81.5	105	2	C72580	hypothetical prote
705	23	85.2	1345	2	A87102	probable SpoIIIE-f	778	22	81.5	107	2	T04306	RikIa protein - ri
706	23	85.2	1385	2	S61236	major capsid prote	779	22	81.5	107	2	T04305	RikIb protein - ri
707	23	85.2	1469	2	B36665	slit protein 2 pre	780	22	81.5	107	2	H81544	hypothetical prote
708	23	85.2	1479	2	T42710	mannose receptor,	781	22	81.5	108	1	NLGA0A	abdominal ganglion
709	23	85.2	1480	2	A36665	slit protein 1 pre	782	22	81.5	108	2	S74608	hypothetical prote
710	23	85.2	1487	2	T02850	hypothetical prote	783	22	81.5	109	2	S51126	genome polyprotein
711	23	85.2	1532	2	A12622	collagen alpha 1(X	784	22	81.5	109	2	F71192	hypothetical prote
712	23	85.2	1538	2	H70846	hypothetical glyci	785	22	81.5	109	2	T27861	hypothetical prote
713	23	85.2	2561	2	T24864	hypothetical prote	786	22	81.5	109	2	T04144	DNA binding protei
714	23	85.2	3436	2	S55659	tegument protein 6	787	22	81.5	111	1	W0BP67	gene 0.6B protein
715	23	85.2	4302	2	J38971	polycystic kidney	788	22	81.5	111	2	S41504	12RNPI protein - S
716	23	85.2	5107	2	T29144	partial CDS - Caen	789	22	81.5	112	2	S51122	genome polyprotein
717	22	81.5	38	2	S55678	calcium-dependent	790	22	81.5	112	2	S51129	genome polyprotein
718	22	81.5	43	2	B41711	defensin B - beetl	791	22	81.5	112	2	T36440	hypothetical prote
719	22	81.5	51	2	AF0260	hypothetical phage	792	22	81.5	113	2	A72380	nitrogen regulator
720	22	81.5	59	2	AF3247	conserved hypothet	793	22	81.5	113	2	S72784	hypothetical prote
721	22	81.5	62	2	A12652	hypothetical prote	794	22	81.5	116	2	B71060	hypothetical prote
722	22	81.5	64	2	B82778	hypothetical prote	795	22	81.5	116	2	B90796	hypothetical prote
723	22	81.5	64	2	B72740	probable-ribonucle	796	22	81.5	116	2	E85605	hypothetical prote
724	22	81.5	66	2	T36809	hypothetical prote	797	22	81.5	116	2	F85957	unknown protein en
725	22	81.5	66	2	T29847	hypothetical prote	798	22	81.5	116	2	D91112	hypothetical prote
726	22	81.5	69	2	D90905	hypothetical prote	799	22	81.5	117	2	C87594	hypothetical prote
727	22	81.5	70	2	B85695	hypothetical prote	800	22	81.5	117	2	H72706	hypothetical prote
728	22	81.5	70	2	B64863	hypothetical prote	801	22	81.5	120	1	W4WLEB	E4 protein - bovin
729	22	81.5	72	2	I53411	acetylcholinestera	802	22	81.5	120	2	E87594	hypothetical prote
730	22	81.5	72	2	B41573	acetylcholinestera	803	22	81.5	122	2	H70370	holo-lacyl-carrier
731	22	81.5	76	2	H81214	hypothetical prote	804	22	81.5	122	2	D72756	hypothetical prote
732	22	81.5	76	2	A39034	neurogranin - bovi	805	22	81.5	123	2	H81130	transcription regu
733	22	81.5	77	2	F87395	conserved hypothet	806	22	81.5	125	2	T00983	cytochrome b6-f co
734	22	81.5	78	1	E69549	conserved hypothet	807	22	81.5	127	2	T22292	hypothetical prote
735	22	81.5	78	2	I47043	neurogranin KC3 fi	808	22	81.5	129	2	T04230	hypothetical prote
736	22	81.5	79	2	AGI080	ribosomal protein	809	22	81.5	130	1	WMBEH3	UL33 protein - hum
737	22	81.5	79	2	AH1437	ribosomal protein	810	22	81.5	130	2	D95965	hypothetical prote
738	22	81.5	79	2	T34777	hypothetical prote	811	22	81.5	131	2	A43980	neurophysin 2 [val
739	22	81.5	81	2	S66013	ribosomal protein	812	22	81.5	131	2	H72246	ribosomal protein
740	22	81.5	84	2	H96495	unknown protein, 4	813	22	81.5	132	2	E72727	probable ribosomal
741	22	81.5	84	2	F72497	hypothetical prote	814	22	81.5	132	2	H87626	hypothetical prote
742	22	81.5	86	2	F69013	hypothetical prote	815	22	81.5	132	2	A82706	hypothetical prote
743	22	81.5	88	2	S77431	hypothetical prote	816	22	81.5	133	2	H72459	probable ribosomal
744	22	81.5	88	2	T01423	hypothetical prote	817	22	81.5	133	2	T04126	V0 protein - Misca
745	22	81.5	89	1	H64116	ribosomal protein	818	22	81.5	134	2	J01355	preprotein translo
746	22	81.5	89	2	T05524	probable cytochrom	819	22	81.5	134	2	A87677	hypothetical prote
747	22	81.5	89	2	AE3314	soluble lytic mure	820	22	81.5	135	1	R5SP16	ribosomal protein
748	22	81.5	90	2	D85845	probable lysis pro	821	22	81.5	135	2	S31385	hypothetical 15.2K
749	22	81.5	93	2	B88449	protein C45G9.12 [	822	22	81.5	135	2	G70259	hypothetical prote
750	22	81.5	93	2	A32327	lampbrush loop-ass	823	22	81.5	135	2	G70816	probable cold choc
751	22	81.5	94	2	A70752	probable rpmB prot	824	22	81.5	135	2	I38917	homeobox, D424 rep
752	22	81.5	94	2	S28802	hypothetical prote	825	22	81.5	136	2	S36304	T-cell receptor de
753	22	81.5	94	2	T46349	hypothetical prote	826	22	81.5	136	2	AB3058	hypothetical prote
754	22	81.5	95	2	S33811	cpgi protein - rat	827	22	81.5	136	2	T45339	small cold-shock p
755	22	81.5	95	2	T45336	hypothetical prote	828	22	81.5	136	2	T12140	legumin - fava bea
756	22	81.5	96	2	G84240	hypothetical prote	829	22	81.5	137	2	C87332	hypothetical prote
757	22	81.5	96	2	JQ2011	hypothetical ilk p	830	22	81.5	137	2	H86410	protein F3M18.8 fi
758	22	81.5	97	2	AD2171	virulence-associat	831	22	81.5	137	2	T51202	hypothetical prote
759	22	81.5	97	2	G75391	hypothetical prote	832	22	81.5	138	2	AC2255	two-component resp

keratin, feather -	hypothetical prote	calcium-binding pr	hypothetical prote	collagen alpha 2(I	probable transcrip	probable whiB3 pro	hypothetical prote	hypothetical prote	hypothetical prote	probable holin pro	hypothetical prote	hypothetical prote	RikIa protein - ri	RikIb protein - ri	hypothetical prote	abdominal ganglion	hypothetical prote	genome polyprotein	hypothetical prote	hypothetical prote	gene 0.6B protein	12RNPI protein - S	genome polyprotein	genome polyprotein	hypothetical prote	nitrogen regulator	hypothetical prote	hypothetical prote	hypothetical prote	E4 protein - bovin	hypothetical prote	holo-lacyl-carrier	transcription regu	cytochrome b6-f co	hypothetical prote	hypothetical prote	UL33 protein - hum	hypothetical prote	neurophysin 2 [val	ribosomal protein	probable ribosomal	hypothetical prote	hypothetical prote	probable ribosomal	V0 protein - Misca	preprotein translo	hypothetical prote	ribosomal protein	hypothetical 15.2K	hypothetical prote	probable cold choc	homeobox, D424 rep	T-cell receptor de	hypothetical prote	small cold-shock p	legumin - fava bea	hypothetical prote	protein F3M18.8 fi	hypothetical prote	two-component resp
--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	-------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	-------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	-------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------

833	22	81.5	138	2	D83639	hypothetical prote	906	22	81.5	172	1	B69278	NADH oxidase (noxC
834	22	81.5	138	2	G87277	hypothetical prote	907	22	81.5	172	2	T36713	hypothetical prote
835	22	81.5	139	2	C72500	hypothetical prote	908	22	81.5	172	2	AB2706	conserved hypotet
836	22	81.5	141	2	G87702	conserved hypotet	909	22	81.5	172	2	B97488	hypothetical prote
837	22	81.5	141	2	F70847	hypothetical prote	910	22	81.5	172	2	A70864	hypothetical prote
838	22	81.5	141	2	S09804	hypothetical prote	911	22	81.5	173	2	A70864	4-hydroxyphenylace
839	22	81.5	141	2	S83079	hypothetical prote	912	22	81.5	173	2	AG3340	hypothetical prote
840	22	81.5	142	2	H86603	hypothetical prote	913	22	81.5	174	1	GI1885	conserved hypotet
841	22	81.5	142	2	A72019	hypothetical prote	914	22	81.5	174	2	D72525	hypothetical prote
842	22	81.5	143	2	S52595	conserved hypotet	915	22	81.5	174	2	T52600	hypothetical prote
843	22	81.5	144	2	F70175	ferredoxin [2Fe-2S	916	22	81.5	175	2	A70952	squamosa promoter
844	22	81.5	144	2	R82455	probable acetyltra	917	22	81.5	175	2	F98824	hypothetical prote
845	22	81.5	144	2	PQ0286	polyprotein - coxs	918	22	81.5	176	2	S39963	hypothetical prote
846	22	81.5	145	2	PQ0287	polyprotein - coxs	919	22	81.5	176	2	C87614	hypothetical prote
847	22	81.5	145	2	A87574	hypothetical prote	920	22	81.5	177	2	B72580	hypothetical prote
848	22	81.5	145	2	H87436	hypothetical prote	921	22	81.5	177	2	MXR12	nonstructural prot
849	22	81.5	145	2	E97206	transcription regu	922	22	81.5	178	1	S00528	vega protein - Myx
850	22	81.5	146	2	H72675	hypothetical prote	923	22	81.5	178	2	T05094	hypothetical prote
851	22	81.5	146	2	S53712	nitric-oxide reduc	924	22	81.5	178	2	C87553	hypothetical prote
852	22	81.5	146	2	E84305	hypothetical prote	925	22	81.5	179	2	D86757	hypothetical prote
853	22	81.5	147	2	S42421	hypothetical prote	926	22	81.5	181	2	T52607	squamosa promoter
854	22	81.5	148	2	D90407	single-stranded DN	927	22	81.5	182	2	G75538	conserved hypotet
855	22	81.5	149	2	AI3264	universal stress p	928	22	81.5	182	2	C82574	phage-related lyso
856	22	81.5	150	2	D72670	hypothetical prote	929	22	81.5	183	2	A86373	protein T23E23.2 [
857	22	81.5	150	2	B98044	ABC transporter, c	930	22	81.5	184	2	T52628	splicing factor RS
858	22	81.5	151	2	H83231	conserved hypotet	931	22	81.5	184	2	C96506	T12C22.1 [imported
859	22	81.5	151	2	H95394	probable transcrip	932	22	81.5	184	2	B87708	hypothetical prote
860	22	81.5	152	2	AI0448	probable ribonucle	933	22	81.5	184	2	T27050	hypothetical prote
861	22	81.5	152	2	AH3013	tolR protein (impo	934	22	81.5	184	2	A61196	genome polyprotein
862	22	81.5	152	2	G98270	tolR protein (tolR	935	22	81.5	185	2	B75586	hypothetical prote
863	22	81.5	152	2	H82578	hypothetical prote	936	22	81.5	185	2	AI2726	urease accessory p
864	22	81.5	153	2	H82676	conserved hypotet	937	22	81.5	185	2	F97690	hypothetical prote
865	22	81.5	154	2	S75674	hypothetical prote	938	22	81.5	186	2	AC1650	hypothetical prote
866	22	81.5	154	2	T35022	probable transcrip	939	22	81.5	187	2	T10073	amine dehydrogenas
867	22	81.5	154	2	R70508	hypothetical prote	940	22	81.5	187	2	T51584	splicing factor 9G
868	22	81.5	155	2	E72460	hypothetical prote	941	22	81.5	188	2	C87341	conserved hypotet
869	22	81.5	156	2	G83199	conserved hypotet	942	22	81.5	188	2	E97508	hypothetical prote
870	22	81.5	157	2	E97424	hypothetical 18.7K	943	22	81.5	189	2	A86369	hypothetical prote
871	22	81.5	157	2	AD2642	conserved hypotet	944	22	81.5	189	2	A48834	basic fibroblast g
872	22	81.5	157	2	C90367	hypothetical prote	945	22	81.5	191	2	H71370	hypothetical prote
873	22	81.5	157	2	R75530	hypothetical prote	946	22	81.5	192	1	ASLJND	vif protein - huma
874	22	81.5	158	2	S76901	hypothetical prote	947	22	81.5	192	2	S42944	viral infectivity
875	22	81.5	158	2	S77877	hypothetical prote	948	22	81.5	192	2	S42993	viral infectivity
876	22	81.5	158	2	T29432	hypothetical prote	949	22	81.5	192	2	E82542	two-component syst
877	22	81.5	159	2	B72456	hypothetical prote	950	22	81.5	192	2	A72646	hypothetical prote
878	22	81.5	159	2	H87539	hypothetical prote	951	22	81.5	193	2	A70583	probable resolvase
879	22	81.5	159	2	JQ0136	conserved hypotet	952	22	81.5	193	2	E83864	thioredoxin BH1717
880	22	81.5	160	2	AG3460	hypothetical 17.8K	953	22	81.5	194	2	S06067	nonstructural prot
881	22	81.5	160	2	AF3104	hypothetical cytos	954	22	81.5	194	2	F97691	hypothetical prote
882	22	81.5	160	2	E98182	conserved hypotet	955	22	81.5	195	1	SAVLDS	delta large antige
883	22	81.5	161	2	S14480	hypothetical prote	956	22	81.5	195	2	E27733	nifQ protein - Azo
884	22	81.5	162	2	T36630	arginine-vasotocin	957	22	81.5	196	2	D69278	3-dehydroquinat d
885	22	81.5	162	2	AG2685	probable secreted	958	22	81.5	196	2	JC4257	beta A4-crystallin
886	22	81.5	162	2	D97467	acetyltransferase	959	22	81.5	196	2	S08677	hypothetical prote
887	22	81.5	166	2	H75163	hypothetical prote	960	22	81.5	197	2	S25106	hypothetical prote
888	22	81.5	166	2	C72734	hypothetical prote	961	22	81.5	197	2	T36696	capsid protein - b
889	22	81.5	166	2	T18691	hypothetical prote	962	22	81.5	197	2	S74830	probable regulator
890	22	81.5	167	2	H70815	probable MoaC-2 pr	963	22	81.5	197	2	D75315	hypothetical prote
891	22	81.5	167	2	E87339	hypothetical prote	964	22	81.5	198	1	FOAD72	rif2 protein - Dei
892	22	81.5	167	2	S73043	hypothetical prote	965	22	81.5	200	1	VCVQBY	major core protein
893	22	81.5	167	2	T52599	squamosa promoter	966	22	81.5	200	1	VCVQC3	coat protein - bar
894	22	81.5	168	2	AE2385	hypothetical prote	967	22	81.5	200	2	E87198	hypothetical prote
895	22	81.5	168	2	T33564	hypothetical prote	968	22	81.5	201	2	A43731	Achaete-scute comp
896	22	81.5	169	2	S72818	heat shock protein	969	22	81.5	201	2	S23324	gene achaeate prote
897	22	81.5	169	2	G72559	hypothetical prote	970	22	81.5	201	2	G86206	hypothetical prote
898	22	81.5	169	2	I57953	peripheral-type be	971	22	81.5	201	2	T49792	hypothetical prote
899	22	81.5	169	2	A53405	peripheral-type be	972	22	81.5	203	2	G70501	probable 3-methylp
900	22	81.5	169	2	JC1393	benzodiazepine rec	973	22	81.5	203	2	E70641	hypothetical prote
901	22	81.5	169	2	G69300	conserved hypotet	974	22	81.5	203	2	H71665	probable ribosomal
902	22	81.5	169	2	T35087	probable transcrip	975	22	81.5	203	2	T24537	hypothetical prote
903	22	81.5	169	2	C75595	hypothetical prote	976	22	81.5	203	2	D97816	50S ribosomal prot
904	22	81.5	170	2	AI3322	hypothetical prote	977	22	81.5	203	2	C95350	hypothetical prote
905	22	81.5	171	2	AB3489	hypothetical membr	978	22	81.5	204	2	T51104	dTDP-4-keto-6-deox

979 22 81.5 204 2 C35878  
980 22 81.5 204 2 T32815  
981 22 81.5 204 2 E88044  
982 22 81.5 204 2 T37145  
983 22 81.5 204 2 S62203  
984 22 81.5 204 2 H87358  
985 22 81.5 204 2 G35276  
986 22 81.5 205 2 C72531  
987 22 81.5 205 2 JU0223  
988 22 81.5 206 1 NBHUIB  
989 22 81.5 206 1 E43670  
990 22 81.5 206 2 T07797  
991 22 81.5 206 2 AC3478  
992 22 81.5 207 2 H5618  
993 22 81.5 207 2 A98123  
994 22 81.5 207 2 H85967  
995 22 81.5 207 2 D65095  
996 22 81.5 208 1 D70764  
997 22 81.5 208 2 AF1504  
998 22 81.5 208 2 E64357  
999 22 81.5 208 2 JN0638  
1000 22 81.5 208 2 T21689

ALIGNMENTS

RESULT 1  
S29829  
N:Alcarnate names: scylliorhinine Z3  
C:Species: Scylliorhinus canicula (smaller spotted catshark, smaller spotted dogfish)  
C:Date: 19-Mar-1997 #sequence revision 19-Mar-1997 #text\_change 09-Jul-2004  
C:Accession: S29829; A41269; S12284  
R:Kouach, M.; Jaquinod, M.; Belaiche, D.; Sautiere, P.; van Dorsselaer, A.; Chevaillier, R.  
Biochim. Biophys. Acta 1162, 99-104, 1993  
A:Title: A corrected primary structure for dog-fish Scylliorhinus caniculus protamine Z3  
A:Reference number: S29829; MUID:93192312; PMID:8448201  
A:Accession: S29829  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-37 <KOU>  
A:Cross-references: UNIPROT:P30258; UNIPARC:UPI00001323AF  
R:Sautiere, P.; Briand, G.; Gusse, M.; Chevaillier, P.  
Eur. J. Biochem. 119, 251-255, 1981  
A:Title: Primary structure of a protamine isolated from the sperm nuclei of the dog-fish  
A:Reference number: A41269; MUID:82072694; PMID:7198042  
A:Accession: A41269  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-31 <SAU>  
A:Cross-references: UNIPARC:UPI0000177209  
R:Gusse, M.; Sautiere, P.; Chauviere, M.; Chevaillier, P.  
Biochim. Biophys. Acta 746, 93-98, 1983  
A:Title: Extraction, purification and characterization of the sperm protamines of the dogfish  
A:Reference number: S01463; MUID:84000513; PMID:6615852  
A:Accession: S12284  
A:Molecule type: protein  
A:Residues: 1-5 <GUS>  
A:Cross-references: UNIPARC:UPI000017720A  
C:Superfamily: protamine Y2  
C:Keywords: DNA binding; nucleus; sperm; testis

Query Match 100.0%; Score 27; DB 2; Length 37;  
Best Local Similarity 100.0%; Pred. No. 47;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
Db 30 GRRGR 34

RESULT 2

C58208  
protamine II-2 - painted turtle  
C:Species: Chrysemys picta (painted turtle)  
C:Date: 08-Nov-1996 #sequence\_revision 08-Nov-1996 #text\_change 09-Jul-2004  
C:Accession: C58208  
R:Hunt, J.G.; Kasinsky, H.E.; Elsev, R.M.; Wright, C.L.; Rice, P.; Bell, J.E.; Sharp, D.  
J. Biol. Chem. 271, 23547-23557, 1996  
A:Title: Protamines of reptiles.  
A:Reference number: A58208; MUID:96394458; PMID:8798564  
A:Accession: C58208  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-45 <HUN>  
A:Cross-references: UNIPROT:Q7LZB1; UNIPARC:UPI00001771F2  
C:Superfamily: sperm histone

Query Match 100.0%; Score 27; DB 2; Length 45;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
Db 12 GRRGR 16

RESULT 3

D58208  
protamine II-3 - painted turtle  
C:Species: Chrysemys picta (painted turtle)  
C:Date: 08-Nov-1996 #sequence\_revision 08-Nov-1996 #text\_change 09-Jul-2004  
C:Accession: D58208  
R:Hunt, J.G.; Kasinsky, H.E.; Elsev, R.M.; Wright, C.L.; Rice, P.; Bell, J.E.; Sharp, D.  
J. Biol. Chem. 271, 23547-23557, 1996  
A:Title: Protamines of reptiles.  
A:Reference number: A58208; MUID:96394458; PMID:8798564  
A:Accession: D58208  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-45 <HUN>  
A:Cross-references: UNIPROT:Q7LZB0; UNIPARC:UPI00001771F3  
C:Superfamily: sperm histone

Query Match 100.0%; Score 27; DB 2; Length 45;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
Db 12 GRRGR 16

RESULT 4

B58208  
protamine II-1 - painted turtle  
C:Species: Chrysemys picta (painted turtle)  
C:Date: 08-Nov-1996 #sequence\_revision 08-Nov-1996 #text\_change 09-Jul-2004  
C:Accession: B58208  
R:Hunt, J.G.; Kasinsky, H.E.; Elsev, R.M.; Wright, C.L.; Rice, P.; Bell, J.E.; Sharp, D.  
J. Biol. Chem. 271, 23547-23557, 1996  
A:Title: Protamines of reptiles.  
A:Reference number: A58208; MUID:96394458; PMID:8798564  
A:Accession: B58208  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-45 <HUN>  
A:Cross-references: UNIPROT:Q7LZB5; UNIPARC:UPI00001771EE  
C:Superfamily: sperm histone

Query Match 100.0%; Score 27; DB 2; Length 45;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5



```

Db      12 GRGR 16
|||||
RESULT 5
F58208
protamine II-5 - painted turtle
C:Species: Chrysemys picta (painted turtle)
C>Date: 08-Nov-1996 #sequence_revision 08-Nov-1996 #text_change 09-Jul-2004
C:Accession: F58208
R:Hunt, J.G.; Kasinsky, H.E.; Elsey, R.M.; Wright, C.L.; Rice, P.; Bell, J.E.; Sharp, D.
J. Biol. Chem. 271, 23547-23557, 1996
A:Title: Protamines of reptiles,
A:Reference number: A58208; MUID:96394458; PMID:8798564
A:Accession: F58208
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-47 <HUN>
A:Cross-references: UNIPROT:Q7LZA7; UNIPARC:UPI00001771F6
C:Superfamily: sperm histone

Query Match      100.0%; Score 27; DB 2; Length 47;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GRGR 5
Db      12 GRGR 16
|||||
RESULT 6
C61510
ribosomal protein S18, cytosolic - African clawed frog (fragment)
C:Species: Xenopus laevis (African clawed frog)
C>Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 09-Jul-2004
C:Accession: C61510
R:Grossberger, D.; Plainik, M.; Marcuz, A.
Comp. Biochem. Physiol. B 98, 127-133, 1991
A:Title: Ribosomal and chromosomal protein cDNA clones of Xenopus laevis thymus isolated
A:Reference number: A61510; MUID:91284580; PMID:2060276
A:Accession: C61510
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-51 <GRO>
A:Cross-references: UNIPROT:Q7LZL4; UNIPARC:UPI00001772A1
C:Superfamily: ribosomal protein S13/S18

Query Match      100.0%; Score 27; DB 2; Length 51;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GRGR 5
Db      39 GRGR 43
|||||
RESULT 7
I40314
filamentous hemagglutinin antigen - Bordetella pertussis (fragment)
C:Species: Bordetella pertussis
C>Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 09-Jul-2004
C:Accession: I40314
R:Mattei, D.; Pichot, F.; Bellalou, J.; Mercereau-Puijalon, O.; Ullmann, A.
FEMS Microbiol. Lett. 37, 73-77, 1986
A:Title: Molecular cloning of a coding sequence of Bordetella pertussis filamentous hema
A:Reference number: I40314
A:Accession: I40314
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-54 <RES>
A:Cross-references: UNIPROT:Q45362; UNIPARC:UPI00000B78AE; GB:M35274; NID:g144043; PIDN:

Query Match      100.0%; Score 27; DB 2; Length 54;

```

```

Best Local Similarity 100.0%; Pred. No. 65;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GRGR 5
Db      28 GRGR 32
|||||
RESULT 8
S34045
protamine - North American opossum
C:Species: Didelphis virginiana, Didelphis marsupialis virginiana (North American opossum)
C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Jul-2004
C:Accession: S34045
R:Winkfein, R.J.; Nishikawa, S.; Connor, W.; Dixon, G.H.
Eur. J. Biochem. 215, 63-72, 1993
A:Title: Characterization of a marsupial sperm protamine gene and its transcripts from t
A:Reference number: S34045; MUID:93345500; PMID:8344286
A:Accession: S34045
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-58 <WIN>
A:Cross-references: UNIPROT:P35305; UNIPARC:UPI000016C416; EMBL:X74044; NID:g407062; PIDN:
C:Superfamily: sperm histone
C:Keywords: DNA binding; nucleus

Query Match      100.0%; Score 27; DB 2; Length 58;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GRGR 5
Db      38 GRGR 42
|||||
RESULT 9
A58208
protamine I-1 - painted turtle
C:Species: Chrysemys picta (painted turtle)
C>Date: 08-Nov-1996 #sequence_revision 08-Nov-1996 #text_change 09-Jul-2004
C:Accession: A58208
R:Hunt, J.G.; Kasinsky, H.E.; Elsey, R.M.; Wright, C.L.; Rice, P.; Bell, J.E.; Sharp, D.
J. Biol. Chem. 271, 23547-23557, 1996
A:Title: Protamines of reptiles.
A:Reference number: A58208; MUID:96394458; PMID:8798564
A:Accession: A58208
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-58 <HUN>
A:Cross-references: UNIPROT:Q7LZB2; UNIPARC:UPI00001771F1
C:Superfamily: sperm histone

Query Match      100.0%; Score 27; DB 2; Length 58;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GRGR 5
Db      24 GRGR 28
|||||
RESULT 10
D90140
hypothetical protein SSO5027 [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C:Accession: D90140
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-y
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.
arrett, R.A.; Ragan, M.A.; Sengen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139

```

A:Accession: D90140  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-59 <KUR>  
A:Cross-references: UNIPROT:Q981D3; UNIPARC:UPI0000064164; GB:AE006641; NID:gi3813141; B  
A:Gene: SS05027

Query Match 100.0%; Score 27; DB 2; Length 59;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
DB 48 GRRGR 52

RESULT 11  
S15538  
homeotic protein Hox A6 - human (fragment)  
N:Alternate names: homeotic protein Hox 1B  
C:Species: Homo sapiens (man)  
C:Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 05-Oct-2004  
C:Accession: S15538  
R:Boncinelli, E.; Acampora, D.; Pannese, M.; d'Esposito, M.; Somma, R.; Gaudino, G.; Stc  
Genome 31, 745-756, 1989  
A:Title: Organization of human class I homeobox genes.  
A:Reference number: S15036; MUID:90215256; PMID:2576652  
A:Accession: S15538  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-66 <BON>  
A:Cross-references: UNIPROT:P31267; UNIPARC:UPI000017A31F  
C:Genetics:  
A:Gene: GDB:H0XA6  
A:Cross-references: GDB:120648; OMIM:142951  
A:Map position: 7p15.3-7p15.3  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F:2-58/Domain: homeobox homology <HOX>

Query Match 100.0%; Score 27; DB 2; Length 66;  
Best Local Similarity 100.0%; Pred. No. 77;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
DB 1 GRRGR 5

RESULT 12  
E45186  
homeotic protein Ghox 4.7 - chicken (fragment)  
C:Species: Gallus gallus (chicken)  
C:Date: 10-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 05-Oct-2004  
C:Accession: E45186  
R:Mackem, S.; Mahon, K.A.  
Development 112, 791-806, 1991  
A:Title: Ghox 4.7: a chick homeobox gene expressed primarily in limb buds with limb-type  
A:Reference number: A45186; MUID:92037185; PMID:1682126  
A:Accession: E45186  
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra  
A:Molecule type: mRNA  
A:Residues: 1-70 <MAC>  
A:Cross-references: UNIPARC:UPI00000FB166  
A:Note: sequence extracted from NCBI backbone (NCBIP:63340)  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F:21-70/Domain: homeobox homology (fragment) <HOX>

Query Match 100.0%; Score 27; DB 2; Length 70;  
Best Local Similarity 100.0%; Pred. No. 81;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5

Db 20 GRRGR 24  
|||||

RESULT 13  
S38675  
hypothetical protein - Streptomyces lividans (fragment)  
C:Species: Streptomyces lividans  
C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 09-Jul-2004  
C:Accession: S38675  
R:Nussbaumer, B.; Wohlleben, W.  
submitted to the EMBL Data Library, November 1993  
A:Description: Identification isolation and sequencing of the recA gene of streptomyces  
A:Reference number: S38674  
A:Accession: S38675  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-75 <NUS>  
A:Cross-references: UNIPROT:Q54391; UNIPARC:UPI00000B90FA; EMBL:X76076; NID:g416192; PIDN

Query Match 100.0%; Score 27; DB 2; Length 75;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
DB 22 GRRGR 26  
|||||

RESULT 14  
T30353  
hypothetical protein ORF6 - Lymantria dispar nuclear polyhedrosis virus  
C:Species: Lymantria dispar nuclear polyhedrosis virus, LdMNPV  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T30353  
R:Kuzio, J.; Pearson, M.N.; Harwood, S.H.; Funk, C.J.; Evans, J.T.; Slavicek, J.M.; Rohrn  
Virology 253, 17-34, 1999  
A:Title: Sequence and analysis of the genome of a baculovirus pathogenic for Lymantria di  
A:Reference number: Z20836; MUID:99124785; PMID:9887315  
A:Accession: T30353  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-80 <KUZ>  
A:Cross-references: UNIPROT:Q9YMW8; UNIPARC:UPI00000F8902; EMBL:AF081810; PIDN:AACT0191.1

Query Match 100.0%; Score 27; DB 2; Length 80;  
Best Local Similarity 100.0%; Pred. No. 90;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
DB 6 GRRGR 10  
|||||

RESULT 15  
B29585  
homeotic protein Hox 2.2 precursor - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 15-Dec-1988 #sequence\_revision 30-Sep-1991 #text\_change 31-Dec-2004  
C:Accession: B29585  
R:Ionai, P.; Arman, E.; Czosnek, H.; Ruddle, F.H.; Blatt, C.  
DNA 6, 409-418, 1987  
A:Title: New murine homeoboxes: structure, chromosomal assignment, and differential expre  
A:Reference number: A29585; MUID:88054465; PMID:2890503  
A:Accession: B29585  
A:Molecule type: DNA  
A:Residues: 1-81 <ION>  
A:Cross-references: UNIPARC:UPI000017A2B5; GB:M18167  
A:Note: the authors translated the codon CAG for residue 69 as Glu  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F:4-60/Domain: homeobox homology <HOX>

Query Match 100.0%; Score 27; DB 2; Length 81;

Best Local Similarity 100.0%; Pred. No. 91;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
|||||  
Db 3 GRGR 7

RESULT 16,  
G70186  
conserved hypothetical protein BB0696 - Lyme disease spirochete  
C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: G70186  
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White  
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,  
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
Nature 390, 580-586, 1997  
A:Authors: Smith, H.O.; Venter, J.C.  
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.  
A:Reference number: A70100; MUID:98065943; PMID:9403685  
A:Accession: G70186  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-82 <KLE>  
A:Cross-references: UNIPROT:O51639; UNIPARC:UPI000005759D; GB:AE001170; GB:AE000783; NID  
A:Experimental source: strain B31  
C:Superfamily: Bacillus conserved hypothetical protein yIqC

Query Match 100.0%; Score 27; DB 1; Length 82;  
Best Local Similarity 100.0%; Pred. No. 92;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
|||||  
Db 52 GRGR 56

RESULT 17  
T42291  
hypothetical protein - phage SPP1  
C:Species: phage SPP1  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: T42291  
R:Alonso, J.C.; Luder, G.; Stiege, A.C.; Chai, S.; Weise, F.; Trautner, T.A.  
Gene 204, 201-212, 1997  
A:Title: The complete nucleotide sequence and functional organization of Bacillus subtil  
A:Reference number: Z2137; MUID:98094274; PMID:9434185  
A:Accession: T42291  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-85 <ALO>  
A:Cross-references: UNIPROT:O48451; UNIPARC:UPI000009C2E3; EMBL:X97918; PIDN:CAA66552.1

Query Match 100.0%; Score 27; DB 2; Length 85;  
Best Local Similarity 100.0%; Pred. No. 95;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
|||||  
Db 4 GRGR 8

RESULT 18  
C27176  
homeotic protein Hox 2.2 - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 05-Oct-2004  
C:Accession: C27176  
R:Hard, C.P.; Fainsod, A.; Ruddle, F.H.  
Genomics 1, 182-195, 1987  
A:Title: Sequence analysis of the murine Hox-2.2, -2.3, and -2.4 homeo boxes: evolutionary  
A:Reference number: A27176; MUID:88085193; PMID:2891608

A:Accession: C27176  
A:Molecule type: DNA  
A:Residues: 1-97 <HAR>  
A:Cross-references: UNIPROT:P09023; UNIPARC:UPI000016CDEF; GB:M18401; NID:g193936; PIDN:  
C:Genetics:  
A:Gene: Hox-2.2  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F;20-76/Domain: homeobox homology <HOX>

Query Match 100.0%; Score 27; DB 2; Length 97;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
|||||  
Db 19 GRGR 23

RESULT 19  
A24779  
homeotic protein m5-4 - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 22-Jul-1987 #sequence\_revision 22-Jul-1987 #text\_change 31-Dec-2004  
C:Accession: A24779  
R:Colberg-Poley, A.M.; Voss, S.D.; Chowdhury, K.; Stewart, C.L.; Wagner, E.F.; Gruss, P.  
Cell 43, 39-45, 1985  
A:Title: Clustered homeo boxes are differentially expressed during murine development.  
A:Reference number: A24779; MUID:86079512; PMID:2416462  
A:Accession: A24779  
A:Molecule type: DNA  
A:Residues: 1-97 <COL>  
A:Cross-references: UNIPROT:P09092; UNIPARC:UPI000016CDDF; GB:M11988; NID:g193900; PIDN:  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F;20-76/Domain: homeobox homology <HOX>

Query Match 100.0%; Score 27; DB 2; Length 97;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
|||||  
Db 19 GRGR 23

RESULT 20  
A47056  
hypothetical protein orfOA cnr-region [imported] - Alcaligenes eutrophus  
C:Species: Alcaligenes eutrophus  
C>Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 20-Jun-2000  
C:Accession: A47056  
R:Liesegang, H.; Lemke, K.; Siddiqui, R.A.; Schlegel, H.G.  
J. Bacteriol. 175, 767-778, 1993  
A:Title: Characterization of the inducible nickel and cobalt resistance determinant cnr  
A:Reference number: A47056; MUID:93139046; PMID:8380802  
A:Contents: CH34, pMOU28  
A:Accession: A47056  
A:Molecule type: nucleic acid  
A>Status: preliminary  
A:Residues: 1-99 <LIE>  
A:Cross-references: UNIPARC:UPI000017AA16  
A>Note: sequence extracted from NCBI backbone (NCBIN:123544, NCBIP:123545)

Query Match 100.0%; Score 27; DB 2; Length 99;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
|||||  
Db 6 GRGR 10

RESULT 21  
A72581

hypothetical protein APE1928 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jul-2004  
C:Accession: A72581  
R:Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K DNA Res. 6, 83-101, 1999  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix strain K1  
A:Reference number: A72450; MUID:99310339; PMID:10382966  
A:Accession: A72581  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-106 <KAW>  
A:Cross-references: UNIPROT:Q9YAL4; UNIPARC:UPI000005E142; DDBJ:AP0000062; NID:g5105244;  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE1928

Query Match 100.0%; Score 27; DB 2; Length 106;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
|||||  
DB 93 GRRGR 97

RESULT 22  
T35057  
hypothetical protein SC4G2.22 SC4G2.22 - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 09-Jul-2004  
C:Accession: T35057  
R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, August 1998  
A:Reference number: Z1566  
A:Accession: T35057  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-112 <SEE>  
A:Cross-references: UNIPROT:Q86694; UNIPARC:UPI00000DADB0; EMBL:AL031371; PIDN:CAA20559.  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SC4G2.22

Query Match 100.0%; Score 27; DB 2; Length 112;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
|||||  
DB 84 GRRGR 88

RESULT 23  
S72797  
probable K+ transport protein tp2 - Mycobacterium leprae  
N:Alternate names: B1549 C2.206 protein  
C:Species: Mycobacterium leprae  
C:Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004  
C:Accession: S72797  
R:Smith, D.R.; Robison, K.  
submitted to the EMBL Data Library, November 1993  
A:Description: Mycobacterium leprae cosmid B1549.  
A:Reference number: S72582  
A:Accession: S72797  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-122 <SMI>  
A:Cross-references: UNIPROT:Q49710; UNIPARC:UPI00000B450D; EMBL:U00014; NID:g466903; PID

Query Match 100.0%; Score 27; DB 2; Length 122;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
|||||  
DB 85 GRRGR 89

RESULT 24  
G81206  
hypothetical protein NMB0369 [imported] - Neisseria meningitidis (strain MC58 serogroup I  
C:Species: Neisseria meningitidis  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 09-Jul-2004  
C:Accession: G81206  
R:Frettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M. Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ver  
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
A:Reference number: A81000; MUID:20175755; PMID:10710307  
A:Accession: G81206  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-129 <TET>  
A:Cross-references: UNIPROT:Q9K125; UNIPARC:UPI00000C4482; GB:AE002393; GB:AE002098; NID:  
A:Experimental source: serogroup B, strain MC58  
C:Genetics:  
A:Gene: NMB0369

Query Match 100.0%; Score 27; DB 2; Length 129;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
|||||  
DB 125 GRRGR 129

RESULT 25  
T46388  
hypothetical protein DKFZp434A1820.1 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 09-Jul-2004  
C:Accession: T46388  
R:Ottenwaelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, January 2000  
A:Reference number: Z23031  
A:Accession: T46388  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-129 <AAA>  
A:Cross-references: UNIPROT:Q9NT41; UNIPARC:UPI000006EFC5; EMBL:AL137545  
A:Experimental source: adult testis; clone DKFZp434A1820  
C:Genetics:  
A:Note: DKFZp434A1820.1

Query Match 100.0%; Score 27; DB 2; Length 129;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
|||||  
DB 55 GRRGR 59

RESULT 26  
B72557  
probable ribosomal protein S11 APE1742 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jul-2004  
C:Accession: B72557  
R:Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Ki

DNA Res. 6, 83-101, 1999  
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum  
 A:Reference number: A72450; MUID:99310339; PMID:10382966  
 A:Accession: B72557  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-131 <KAW>  
 A:Cross-references: UNIPROT:Q9YB55; UNIPARC:UPI0000035E083; DDBJ:AP0000062; NID:G5105244;  
 A:Experimental source: strain K1  
 C:Genetics:  
 A:Gene: APE1742  
 C:Superfamily: ribosomal protein S11/S14

Query Match 100.0%; Score 27; DB 2; Length 131;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
 Db 125 GRGR 129

RESULT 27  
 S47022  
 ribosomal protein S11 - Sulfolobus acidocaldarius  
 C:Species: Sulfolobus acidocaldarius  
 C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 09-Jul-2004  
 A:Accession: S47022  
 R:Langer, D.; Hain, J.; Thuriaux, P.; Zillig, W.  
 submitted to the EMBL Data Library, July 1984  
 A:Description: Similarity of the transcription systems of Eukarya and Archaea.  
 A:Reference number: S47020  
 A:Accession: S47022  
 A:Molecule type: DNA  
 A:Residues: 1-132 <LAN>  
 A:Cross-references: UNIPROT:P39459; UNIPARC:UPI0000134C18; EMBL:X80194; NID:G517286; PID  
 C:Genetics:  
 A:Gene: rps11  
 C:Superfamily: ribosomal protein S11/S14  
 C:Keywords: protein biosynthesis; ribosome

Query Match 100.0%; Score 27; DB 2; Length 132;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
 Db 126 GRGR 130

RESULT 28  
 S75416  
 ribosomal protein S14 - Sulfolobus solfataricus  
 A:Alternate names: protein c04050  
 C:Species: Sulfolobus solfataricus  
 C:Date: 11-Oct-1997 #sequence\_revision 24-Oct-1997 #text\_change 09-Jul-2004  
 A:Accession: S75416  
 R:Sensen, C.W.; Klenk, H.P.; Singh, R.K.; Allard, G.; Chan, C.C.Y.; Liu, Q.Y.; Penny, S.  
 Mol. Microbiol. 22, 175-191, 1996  
 A:Title: Organizational characteristics and information content of an archaeal genome: 1  
 A:Reference number: S73076; MUID:97055432; PMID:8899719  
 A:Accession: S75416  
 A>Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-132 <SEN>  
 A:Cross-references: UNIPROT:P95988; UNIPARC:UPI0000134C19; EMBL:Y08257; NID:G1707772; PI  
 A:Experimental source: strain P2  
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1996  
 C:Genetics:  
 A:Start codon: GTG  
 C:Superfamily: ribosomal protein S11/S14  
 C:Keywords: protein biosynthesis; ribosome

Query Match 100.0%; Score 27; DB 2; Length 132;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
 Db 126 GRGR 130

RESULT 29  
 JC4877  
 phospholipase A2 homolog 3 precursor - himenobu  
 C:Species: Trimeresurus okinawensis (himenobu)  
 C:Date: 15-Aug-1996 #sequence\_revision 14-Feb-1997 #text\_change 16-Aug-2004  
 C:Accession: JC4877  
 R:Nobuhisa, I.; Nakashima, K.; Deshimaru, M.; Ogawa, T.; Shimohigashi, Y.; Fukumaki, Y.;  
 Gene 172, 267-272, 1996  
 A:Title: Accelerated evolution of Trimeresurus okinawensis venom gland phospholipase A2  
 A:Reference number: JC4874; MUID:96269416; PMID:8682315  
 A:Contents: venom gland  
 A:Accession: JC4877  
 A:Molecule type: mRNA  
 A:Residues: 1-137 <NOB>  
 A:Cross-references: UNIPROT:Q92152; UNIPARC:UPI00001310B4; DDBJ:D49389; NID:G1469806; PI  
 C:Comment: This protein probably does not have phospholipase A2 activity.  
 C:Superfamily: Phospholipase A2  
 C:Keywords: venom  
 F:1-16/Domain: signal sequence #status predicted <SIG>  
 F:17-137/Product: phospholipase A2 homolog 3 #status predicted <MAT>  
 F:42-131,44-60,59-111,65-137,66-104,73-97,91-102/Disulfide bonds: #status predicted

Query Match 100.0%; Score 27; DB 1; Length 137;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
 Db 47 GRGR 51

RESULT 30  
 R5BY59  
 ribosomal protein S14.e.A, cytosolic - yeast (Saccharomyces cerevisiae)  
 A:Alternate names: protein YCR031c; ribosomal protein S9  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 09-Jul-2004  
 C:Accession: A02726; A26887; S19443  
 R:Teem, J.L.; Abovich, N.; Kaufer, N.F.; Schwindinger, W.F.; Warner, J.R.; Levy, A.; Woolf  
 , H.; Roebash, M.  
 Nucleic Acids Res. 12, 8295-8312, 1984  
 A:Title: A comparison of yeast ribosomal  
 A:Reference number: A93541; MUID:85062814; PMID:6390341  
 A:Accession: A02726  
 A:Molecule type: DNA  
 A:Residues: 1-137 <TEE>  
 A:Cross-references: UNIPROT:P06367; UNIPARC:UPI0000168BA4  
 R:Larkin, J.C.; Thompson, J.R.; Woolford Jr., J.L.  
 Mol. Cell. Biol. 7, 1764-1775, 1987  
 A:Title: Structure and expression of the Saccharomyces cerevisiae CRY1 gene: a highly cor  
 A:Reference number: A26887; MUID:87257876; PMID:3037334  
 A:Accession: A26887  
 A:Molecule type: DNA  
 A:Residues: 1-137 <LAR>  
 A:Cross-references: UNIPARC:UPI0000168BA4; GB:M16126; NID:G171321; PIDN:AAA34530.1; PID:  
 R:Cederberg, H.; Hohmann, S.; Schaeff-Gerstenschlager, I.; Huse, K.; Zimmermann, F.K.  
 submitted to the Protein Sequence Database, March 1992  
 A:Reference number: S19439  
 A:Accession: S19443  
 A:Molecule type: DNA  
 A:Residues: 1-137 <CED>  
 A:Cross-references: UNIPARC:UPI0000168BA4; EMBL:X59720; GSPDB:GN00003; MIPS:YCR031c  
 C:Genetics:  
 A:Gene: SGD:CRY1; RP59; MIPS:YCR031c

A;Cross-references: SGD:S0000627; MIPS:YCR031c  
A;Map position: 3R  
A;Introns: 3/1  
C;Superfamily: ribosomal protein S11/S14  
C;Keywords: protein biosynthesis; ribosome

Query Match 100.0%; Score 27; DB 1; Length 137;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GRRGR 5  
Db 131 GRRGR 135

RESULT 31  
S30002  
ribosomal protein S14 e, cytosolic - yeast (Kluyveromyces marxianus)  
N;Alternate names: ribosomal protein RP59  
C;Species: Kluyveromyces marxianus  
C;Date: 07-May-1993 #sequence\_revision 07-May-1993 #text\_change 12-Jun-2003  
C;Accession: S30002  
R;Bergkamp-Steffens, G.K.; Hoekstra, R.; Planta, R.J.  
Yeast 8, 903-922, 1992  
A;Title: Structural and putative regulatory sequences of Kluyveromyces ribosomal protein  
A;Reference number: S29999; MUID:93127726; PMID:1481569  
A;Accession: S30002  
A;Molecule type: DNA  
A;Residues: 1-137 <BER>  
A;Cross-references: UNIPARC:UPI0000134CD5; GB:S53438; NID:g263480; PIDN:AAB24899.1; PID:  
A;Experimental source: strain CBS2360  
A;Note: the sequence from Fig. 3 is inconsistent with that from Fig. 2 in having 116-Gly  
C;Genetics:  
A;Gene: RP59  
A;Introns: 4/1  
C;Function:  
A;Pathway: protein biosynthesis  
C;Superfamily: ribosomal protein S11/S14  
C;Keywords: protein biosynthesis; ribosome

Query Match 100.0%; Score 27; DB 2; Length 137;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GRRGR 5  
Db 131 GRRGR 135

RESULT 32  
S22312  
ribosomal protein S14 e, cytosolic - yeast (Kluyveromyces marxianus var. lactis)  
N;Alternate names: ribosomal protein 59  
C;Species: Kluyveromyces marxianus var. lactis, Candida sphaerica  
C;Date: 29-Jan-1993 #sequence\_revision 29-Jan-1993 #text\_change 09-Jul-2004  
C;Accession: S22312  
R;Larson, G.P.; Rossi, J.J.  
Nucleic Acids Res. 19, 4701-4707, 1991  
A;Title: Altered response to growth rate changes in Kluyveromyces lactis versus Saccharo  
A;Reference number: S22312; MUID:91367668; PMID:1891361  
A;Accession: S22312  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-137 <LAR>  
A;Cross-references: UNIPROT:P27069; UNIPARC:UPI0000134CD5; EMBL:X59860; NID:g2896; PIDN:  
A;Experimental source: strain Y1140  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1991  
C;Genetics:  
A;Gene: RP59  
A;Introns: 4/1  
C;Function:  
A;Pathway: protein biosynthesis  
C;Superfamily: ribosomal protein S11/S14

A;Cross-references: SGD:S0000627; MIPS:YCR031c  
A;Map position: 3R  
A;Introns: 3/1  
C;Superfamily: ribosomal protein S11/S14  
C;Keywords: protein biosynthesis; ribosome

Query Match 100.0%; Score 27; DB 2; Length 137;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GRRGR 5  
Db 131 GRRGR 135

RESULT 33  
D75171  
ssu ribosomal protein s11p (rps11p) PAB0362 - Pyrococcus abyssi (strain Orsay)  
C;Species: Pyrococcus abyssi  
C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jul-2004  
C;Accession: D75171  
R;anonymous, Genoscope  
submitted to the EMBL Data Library, July 1999  
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome struc  
A;Reference number: A75001  
A;Accession: D75171  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-137 <KAW>  
A;Cross-references: UNIPROT:P62010; UNIPARC:UPI0000034483; GB:AJ248284; GB:AL096836; NID:  
A;Experimental source: strain Orsay  
C;Genetics:  
A;Gene: rps11p; PAB0362  
C;Superfamily: ribosomal protein S11/S14

Query Match 100.0%; Score 27; DB 2; Length 137;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GRRGR 5  
Db 131 GRRGR 135

RESULT 34  
F71043  
probable ribosomal protein S11 - Pyrococcus horikoshii  
C;Species: Pyrococcus horikoshii  
C;Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 05-Oct-2004  
C;Accession: F71043  
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine  
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,  
DNA Res. 5, 55-76, 1998  
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a  
A;Reference number: A71000; MUID:98344137; PMID:9679194  
A;Accession: F71043  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-137 <KAW>  
A;Cross-references: UNIPROT:P62011; UNIPARC:UPI0000034483; GB:AP000006; NID:g3236133; PII  
A;Experimental source: strain OT3  
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank  
C;Genetics:  
A;Gene: PH1638  
C;Superfamily: ribosomal protein S11/S14

Query Match 100.0%; Score 27; DB 2; Length 137;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GRRGR 5  
Db 131 GRRGR 135

RESULT 35  
S46643

C;Keywords: protein biosynthesis; ribosome

Query Match 100.0%; Score 27; DB 2; Length 137;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GRRGR 5  
Db 131 GRRGR 135

RESULT 33  
D75171  
ssu ribosomal protein s11p (rps11p) PAB0362 - Pyrococcus abyssi (strain Orsay)  
C;Species: Pyrococcus abyssi  
C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jul-2004  
C;Accession: D75171  
R;anonymous, Genoscope  
submitted to the EMBL Data Library, July 1999  
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome struc  
A;Reference number: A75001  
A;Accession: D75171  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-137 <KAW>  
A;Cross-references: UNIPROT:P62010; UNIPARC:UPI0000034483; GB:AJ248284; GB:AL096836; NID:  
A;Experimental source: strain Orsay  
C;Genetics:  
A;Gene: rps11p; PAB0362  
C;Superfamily: ribosomal protein S11/S14

Query Match 100.0%; Score 27; DB 2; Length 137;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GRRGR 5  
Db 131 GRRGR 135

RESULT 34  
F71043  
probable ribosomal protein S11 - Pyrococcus horikoshii  
C;Species: Pyrococcus horikoshii  
C;Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 05-Oct-2004  
C;Accession: F71043  
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine  
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,  
DNA Res. 5, 55-76, 1998  
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a  
A;Reference number: A71000; MUID:98344137; PMID:9679194  
A;Accession: F71043  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-137 <KAW>  
A;Cross-references: UNIPROT:P62011; UNIPARC:UPI0000034483; GB:AP000006; NID:g3236133; PII  
A;Experimental source: strain OT3  
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank  
C;Genetics:  
A;Gene: PH1638  
C;Superfamily: ribosomal protein S11/S14

Query Match 100.0%; Score 27; DB 2; Length 137;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GRRGR 5  
Db 131 GRRGR 135

RESULT 35  
S46643

ribosomal protein S14.e.B, cytosolic - yeast (Saccharomyces cerevisiae)  
 N;Alternate names: protein J0353; protein YUL191w; ribosomal protein rp59  
 C;Species: Saccharomyces cerevisiae  
 C;Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 09-Jul-2004  
 C;Accession: S46643; S56974; S56978; S52592  
 R;Purnelle, B.; Coster, F.; Goffeau, A.  
 Yeast 10, 1235-1249, 1994  
 A;Title: The sequence of a 36 kb segment on the left arm of yeast chromosome X identifies  
 a gene ACO1 and two homologues to chromosome III genes.  
 A;Reference number: S46621; MUID:95274326; PMID:7754713  
 A;Accession: S46643  
 A;Status: nucleic acid sequence not shown  
 A;Molecule type: DNA  
 A;Residues: 1-138 <PUR>  
 A;Cross-references: UNIPROT:P39516; UNIPARC:UPI0000168356; EMBL:X77688  
 R;Obermaier, B.; Piravandi, E.; Rinke, M.; Domdey, H.  
 submitted to the Protein Sequence Database, September 1995  
 A;Reference number: S56937  
 A;Accession: S56974  
 A;Molecule type: DNA  
 A;Residues: 1-138 <OBE>  
 A;Cross-references: UNIPARC:UPI0000168356; EMBL:Z49466; NID:g1008401; PIDN:CAA89486.1; F  
 R;Purnelle, B.; Coster, F.; Goffeau, A.  
 submitted to the Protein Sequence Database, September 1995  
 A;Reference number: S56977  
 A;Accession: S56978  
 A;Molecule type: DNA  
 A;Residues: 1-138 <PUW>  
 A;Cross-references: UNIPARC:UPI0000168356; EMBL:Z49466; MIPS:YUL191w; NID:g1008401; PIDN  
 R;Paulovich, A.G.; Thompson, J.R.; Larkin, J.C.; Li, Z.; Woolford Jr., J.L.  
 Genetics 135, 719-730, 1993  
 A;Title: Molecular genetics of cryptotolerine resistance in Saccharomyces cerevisiae: ex  
 A;Reference number: S52592; MUID:94123967; PMID:8293976  
 A;Accession: S52592  
 A;Molecule type: DNA  
 A;Residues: 1-138 <PAU>  
 A;Cross-references: UNIPARC:UPI0000168356; EMBL:L12564; NID:g295649; PIDN:AAAL7764.1; PI  
 A;Note: the authors did not translate the codons for residues 29 and 30  
 C;Genetics:  
 A;Gene: SGD:CRY2  
 A;Cross-references: SGD:S0003727; MIPS:YUL191w  
 A;Map position: 10L  
 A;Introns: 4/1  
 C;Function:  
 A;Description: protein biosynthesis  
 A;Superfamily: ribosomal protein S11/S14  
 C;Keywords: protein biosynthesis; ribosome

Query Match 100.0%; Score 27; DB 2; Length 138;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
 Db 132 GRGR 136

RESULT 36  
 T38751  
 40S ribosomal protein S14 subunit - fission yeast (Schizosaccharomyces pombe)  
 C;Species: Schizosaccharomyces pombe  
 C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
 C;Accession: T38751; T39777  
 R;Oliver, K.; Harris, D.; Wood, V.; Barrell, B.G.; Rajandream, M.A.  
 submitted to the EMBL Data Library, September 1996  
 A;Reference number: 221809  
 A;Accession: T38751  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-139 <OLI>  
 A;Cross-references: UNIPROT:O14150; UNIPARC:UPI0000134CF1; EMBL:Z99296; PIDN:CAB16591.1;  
 A;Experimental source: strain 972h-; cosmid c3H5  
 R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Badcock, K.; Churcher, C.M.

submitted to the EMBL Data Library, March 1998  
 A;Reference number: 221879  
 A;Accession: T39777  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-139 <LYN>  
 A;Cross-references: UNIPARC:UPI0000134CF1; EMBL:AL022304; PIDN:CAA18410.1; GSPDB:GN00067,  
 A;Experimental source: strain 972h-; cosmid c18H10  
 C;Genetics: <OLI>  
 A;Gene: SPAC3H5.05C  
 A;Map position: 1  
 C;Genetics: <LYN>  
 A;Gene: SPBC18H10.13  
 A;Map position: 2  
 C;Superfamily: ribosomal protein S11/S14

Query Match 100.0%; Score 27; DB 2; Length 139;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
 Db 133 GRGR 137

RESULT 37  
 A24154  
 85K major surface antigen - Trypanosoma cruzi (fragment)  
 C;Species: Trypanosoma cruzi  
 C;Date: 23-Aug-1987 #sequence\_revision 23-Aug-1987 #text\_change 09-Jul-2004  
 C;Accession: A24154  
 R;Peterson, D.S.; Wrightsman, R.A.; Manning, J.E.  
 Nature 322, 566-568, 1986  
 A;Title: Cloning of a major surface-antigen gene of Trypanosoma cruzi and identification  
 A;Reference number: A24154; MUID:86285016; PMID:2426602  
 A;Accession: A24154  
 A;Molecule type: mRNA  
 A;Residues: 1-139 <PET>  
 A;Cross-references: UNIPROT:Q26948; UNIPARC:UPI00007F44B; GB:X04186; NID:g10657; PIDN:C  
 C;Superfamily: trypanastigote-specific surface antigen  
 C;Keywords: surface antigen

Query Match 100.0%; Score 27; DB 2; Length 139;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
 Db 9 GRGR 13

RESULT 38  
 A30097  
 Ribosomal protein S14 (clone MCH1) - maize  
 C;Species: Zea mays (maize)  
 C;Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 09-Jul-2004  
 C;Accession: A30097  
 R;Larkin, J.C.; Hunsperger, J.P.; Culley, D.; Rubenstein, I.; Silflow, C.D.  
 Genes Dev. 3, 500-509, 1989  
 A;Title: The organization and expression of a maize ribosomal protein gene family.  
 A;Reference number: A91625; MUID:89252825; PMID:2470645  
 A;Accession: A30097  
 A;Status: not compared with conceptual translation  
 A;Molecule type: mRNA  
 A;Residues: 1-149 <LAB>  
 A;Cross-references: UNIPROT:P19950; UNIPARC:UPI0000132EE3  
 C;Superfamily: ribosomal protein S11/S14  
 C;Keywords: protein biosynthesis; ribosome

Query Match 100.0%; Score 27; DB 2; Length 149;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
 Db 143 GRRGR 147

RESULT 39  
 S11667  
 ribosomal protein S14.e - Neurospora crassa  
 N:Alternate names: ribosomal protein crp-2  
 C:Species: Neurospora crassa  
 C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 12-Jun-2003  
 C:Accession: S11667  
 R:Tyler, B.M.; Harrison, K.  
 Nucleic Acids Res. 18, 5759-5765, 1990  
 A:Title: A Neurospora crassa ribosomal protein gene, homologous to yeast CRV1, contains  
 A:Reference number: S11667; MUID:91016912; PMID:1977135  
 A:Accession: S11667  
 A:Molecule type: DNA  
 A:Residues: 1-150 <TYL>  
 A:Cross-references: UNIPARC:UPI0000000AF4; EMBL:X53734  
 A:Note: the authors translated the codon GCC for residue 82 as Gln and AGG for residue 1  
 C:Genetics:  
 A:Gene: crp-2  
 A:Introns: 29/1, 47/2, 101/1  
 C:Superfamily: ribosomal protein S11/S14  
 C:Keywords: protein biosynthesis; ribosome

Query Match 100.0%; Score 27; DB 2; Length 150;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
 Db 144 GRRGR 148

RESULT 40  
 B30097  
 ribosomal protein S14 (clone MCH2) - maize  
 C:Species: Zea mays (maize)  
 C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 09-Jul-2004  
 C:Accession: B30097  
 R:Larkin, J.C.; Hunsperger, J.P.; Culley, D.; Rubenstein, I.; Silflow, C.D.  
 Genes Dev. 3, 500-509, 1989  
 A:Title: The organization and expression of a maize ribosomal protein gene family.  
 A:Reference number: A91625; MUID:89252825; PMID:2470645  
 A:Accession: B30097  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-150 <LAR>  
 A:Cross-references: UNIPROT:P19951; UNIPARC:UPI0000132BE6  
 C:Superfamily: ribosomal protein S11/S14  
 C:Keywords: protein biosynthesis; ribosome

Query Match 100.0%; Score 27; DB 2; Length 150;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
 Db 144 GRRGR 148

RESULT 41  
 D84777  
 40S ribosomal protein S14 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
 C:Accession: D84777  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
 Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: A84420; MUID:20083487; PMID:10617197  
 A:Accession: D84777  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-150 <STO>  
 A:Cross-references: UNIPROT:Q9SIH0; UNIPARC:UPI0000132EE2; GB:AE002093; NID:G4678226; PII  
 C:Genetics:  
 A:Gene: At2g36160  
 A:Map position: 2  
 C:Superfamily: ribosomal protein S11/S14

Query Match 100.0%; Score 27; DB 2; Length 150;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
 Db 144 GRRGR 148

RESULT 42  
 T08441  
 ribosomal protein S14, cytosolic - Arabidopsis thaliana  
 N:Alternate names: protein F2206.40  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004  
 C:Accession: T08441  
 R:Quetier, F.; Purnelle, B.; Boutry, M.; Goffeau, A.; Salanoubat, M.; Mewes, H.W.; Mayer,  
 submitted to the Protein Sequence Database, May 1999  
 A:Reference number: Z16420  
 A:Accession: T08441  
 A:Molecule type: DNA  
 A:Residues: 1-150 <QUE>  
 A:Cross-references: UNIPROT:P42036; UNIPARC:UPI0000132EE8; EMBL:AL050300; GSPDB:GN00061;  
 A:Experimental source: cultivar Columbia; BAC clone F2206  
 C:Genetics:  
 A:Gene: ATSP:F2206.40  
 A:Map position: 3  
 A:Introns: 1/3; 41/3; 57/1; 85/3; 129/1  
 C:Superfamily: ribosomal protein S11/S14  
 C:Keywords: protein biosynthesis; ribosome

Query Match 100.0%; Score 27; DB 2; Length 150;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
 Db 144 GRRGR 148

RESULT 43  
 T07974  
 ribosomal protein S14 - yellow lupine  
 C:Species: Lupinus luteus (yellow lupine)  
 C:Date: 21-May-1999 #sequence\_revision 21-May-1999 #text\_change 09-Jul-2004  
 C:Accession: T07974  
 R:Cherepneva, G.N.; Kusnetsov, V.V.; Oelmueller, R.  
 submitted to the EMBL Data Library, September 1997  
 A:Reference number: Z16261  
 A:Accession: T07974  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-150 <CHE>  
 A:Cross-references: UNIPROT:O22584; UNIPARC:UPI0000134CD7; EMBL:AF026079; NID:G2565339; I  
 C:Genetics:  
 A:Gene: rps14  
 C:Superfamily: ribosomal protein S11/S14  
 C:Keywords: ribosome

Query Match 100.0%; Score 27; DB 2; Length 150;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;



Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
|||||

Db 144 GRGR 148

RESULT 44  
R3RT14  
ribosomal protein S14, cytosolic [similarity] - rat  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 09-Jul-2004  
C;Accession: S14900; S06197  
R;Faz, V.; Chan, Y.L.; Glueck, A.; Wool, I.G.  
Nucleic Acids Res. 17, 9484, 1989  
A;Title: The primary structure of rat ribosomal protein S14.  
A;Reference number: S14900; MUID:90067957; PMID:2587275  
A;Accession: S14900  
A;Molecule type: mRNA  
A;Residues: 1-151 <PAZ>  
A;Cross-references: UNIPROT:P13471; UNIPARC:UPI0000134CEB; EMBL:X15040; NID:g57128; PIDN:  
C;Superfamily: ribosomal protein S11/S14  
C;Keywords: protein biosynthesis; ribosome

Query Match 100.0%; Score 27; DB 1; Length 151;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
|||||

Db 145 GRGR 149

RESULT 45  
R4HY14  
ribosomal protein S14 - Chinese hamster  
C;Species: Cricetus griseus (Chinese hamster)  
C;Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 09-Jul-2004  
C;Accession: A02725; I48124  
R;Rhoads, D.D.; Roufa, D.J.  
Mol. Cell. Biol. 5, 1655-1659, 1985  
A;Title: Emetine resistance of Chinese hamster cells: structures of wild-type and mutant  
A;Reference number: A02725; MUID:85267682; PMID:3839563  
A;Accession: A02725  
A;Molecule type: mRNA  
A;Residues: 1-151 <RHO>  
A;Cross-references: UNIPROT:P06366; UNIPARC:UPI0000003F58; GB:M11241; NID:g191197; PIDN:  
R;Rhoads, D.; Roufa, D.J.  
Mol. Biol. Evol. 8, 503-514, 1991  
A;Title: Molecular evolution of the mammalian ribosomal protein gene.  
A;Reference number: I48124; MUID:92017216; PMID:1921707  
A;Accession: I48124  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-151 <RES>  
A;Cross-references: UNIPARC:UPI0000003F58; GB:M35008; NID:g191199; PIDN:AAA37017.1; PID:  
C;Genetics:  
A;Gene: RPS14  
A;Map position: 2q  
A;Intons: 50/2; 104/2; 130/1  
C;Superfamily: ribosomal protein S11/S14  
C;Keywords: protein biosynthesis; ribosome

Query Match 100.0%; Score 27; DB 1; Length 151;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
|||||

Db 145 GRGR 149

RESULT 46

S43296  
bone morphogenetic protein-related protein (GDF7) - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 20-Oct-1994 #sequence\_revision 10-Nov-1995 #text\_change 09-Jul-2004  
C;Accession: S43296  
R;Storm, E.E.; Huynh, T.V.; Copeland, N.G.; Jenkins, N.A.; Kingsley, D.M.; Lee, S.J.  
Nature 368, 639-643, 1994  
A;Title: Limb alterations in brachypodism mice due to mutations in a new member of the  
A;Reference number: S43294; MUID:94195427; PMID:8145850  
A;Accession: S43296  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-151 <STO>  
A;Cross-references: UNIPROT:P43029; UNIPARC:UPI0000027BA4; GB:U08339; NID:g488465; PIDN:  
C;Superfamily: inhibin

Query Match 100.0%; Score 27; DB 2; Length 151;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
|||||

Db 43 GRGR 47

RESULT 47  
A30815  
ribosomal protein S14 - fruit fly (Drosophila melanogaster)  
C;Species: Drosophila melanogaster  
C;Date: 01-Dec-1989 #sequence\_revision 30-Sep-1991 #text\_change 09-Jul-2004  
C;Accession: A30815  
R;Brown, S.J.; Rhoads, D.D.; Stewart, M.J.; Van Slyke, B.; Chen, I.T.; Johnson, T.K.; Der  
Mol. Cell. Biol. 8, 4314-4321, 1988  
A;Title: Ribosomal protein S14 is encoded by a pair of highly conserved, adjacent genes  
A;Reference number: A30815; MUID:89039859; PMID:3141788  
A;Accession: A30815  
A;Molecule type: DNA  
A;Residues: 1-151 <BRO>  
A;Cross-references: UNIPROT:P14130; UNIPARC:UPI0000000FC4; GB:M21045; NID:g158296; PIDN:  
C;Genetics:  
A;Gene: FlyBase:RpS14A  
A;Cross-references: FlyBase:FBgn0004403  
A;Map position: X  
C;Superfamily: ribosomal protein S11/S14  
C;Keywords: protein biosynthesis; ribosome

Query Match 100.0%; Score 27; DB 2; Length 151;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
|||||

Db 145 GRGR 149

RESULT 48  
A25220  
ribosomal protein S14, cytosolic - human  
C;Species: Homo sapiens (man)  
C;Date: 15-Dec-1988 #sequence\_revision 15-Dec-1988 #text\_change 09-Jul-2004  
C;Accession: A25220; A25899; S68925  
R;Rhoads, D.D.; Dixit, A.; Roufa, D.J.  
Mol. Cell. Biol. 6, 2774-2783, 1986  
A;Title: Primary structure of human ribosomal protein S14 and the gene that encodes it.  
A;Reference number: A25220; MUID:87064583; PMID:3785212  
A;Accession: A25220  
A;Molecule type: DNA  
A;Residues: 1-151 <RHO>  
A;Cross-references: UNIPROT:P06366; UNIPARC:UPI0000003F58; GB:M13641; NID:g333  
R;Chen, I.T.; Dixit, A.; Rhoads, D.D.; Roufa, D.J.  
Proc. Natl. Acad. Sci. U.S.A. 83, 6907-6911, 1986  
A;Title: Homologous ribosomal proteins in bacteria, yeast, and humans.  
A;Reference number: A94124; MUID:86313681; PMID:3529092

A:Accession: A25899  
A:Molecule type: mRNA  
A:Residues: 1-151 <CHE>  
A:Cross-references: UNIPARC:UPI0000003F58; GB:M13934; GB:M13641; NID:g337498; PIDN:AAB59  
R:Vladimirov, S.N.; Ivanov, A.V.; Karpova, G.G.; Musolyamov, A.K.; Egorov, T.A.; Thiede,  
Eur. J. Biochem. 239, 144-149, 1996  
A:Title: Characterization of the human small-ribosomal-subunit proteins by N-terminal an  
A:Reference number: S68911; MUID:96305378; PMID:8706699  
A:Accession: S68925  
A:Molecule type: protein  
A:Residues: 2-20 <VLA>  
A:Cross-references: UNIPARC:UPI00000177294  
C:Genetics:  
A:Gene: GDB:RPS14  
A:Cross-references: GDB:119572; OMIM:130620  
A:Map position: 5q31-5q33  
C:Superfamily: ribosomal protein S11/S14  
C:Keywords: protein biosynthesis; ribosome  
F:2-151/Product: ribosomal protein S14, cytosolic #status experimental <MAT>  
  
Query Match 100.0%; Score 27; DB 2; Length 151;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GRGR 5  
Db 145 GRGR 149  
|||||  
  
RESULT 49  
JB0129  
ribosomal protein S14 - mouse  
N:Alternate names: rps 14  
C:Species: Mus musculus (house mouse)  
C:Date: 03-Jun-1998 #sequence\_revision 10-Jul-1998 #text\_change 09-Jul-2004  
C:Accession: JE0129  
R:Lee, M.; Hwang, I.; Choi, Y.; Baik, M.  
Biosci. Biotechnol. Biochem. 62, 573-574, 1998  
A:Title: Sequence of a cDNA encoding mouse ribosomal protein S14.  
A:Reference number: JE0129; MUID:98233282; PMID:9571789  
A:Accession: JE0129  
A:Molecule type: mRNA  
A:Residues: 1-151 <LEE>  
A:Cross-references: UNIPROT:O70569; UNIPARC:UPI0000177293; GB:Y08307  
A:Note: the authors translated the codon CRT for residue 47 as Ser  
C:Superfamily: ribosomal protein S11/S14  
C:Keywords: mammary gland  
  
Query Match 100.0%; Score 27; DB 2; Length 151;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GRGR 5  
Db 145 GRGR 149  
|||||  
  
RESULT 50  
R38T18  
ribosomal protein S18, cytosolic [validated] - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 09-Jul-2004  
C:Accession: JH0419  
R:Chan, Y.L.; Paz, V.; Wool, I.G.  
Biochem. Biophys. Res. Commun. 178, 1212-1218, 1991  
A:Title: The primary structure of rat ribosomal protein S18.  
A:Reference number: JH0419; MUID:91337062; PMID:1872840  
A:Accession: JH0419  
A:Molecule type: mRNA  
A:Residues: 1-152 <CHA>  
A:Cross-references: UNIPROT:P25232; UNIPARC:UPI0000040B5; GB:X57529; NID:9433446; PIDN:  
A:Note: the protein is designated as ribosomal protein S18 according to comigration anal  
C:Superfamily: ribosomal protein S13/S18

C:Keywords: protein biosynthesis; ribosome  
  
Query Match 100.0%; Score 27; DB 1; Length 152;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GRGR 5  
Db 140 GRGR 144  
|||||  
  
RESULT 51  
S30393  
ribosomal protein S18, cytosolic - human  
C:Species: Homo sapiens (man)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: S30393; S68935  
R:Chassin, D.; Bellet, D.; Koman, A.  
Nucleic Acids Res. 21, 745, 1993  
A:Title: The human homolog of ribosomal protein S18.  
A:Reference number: S30393; MUID:93181276; PMID:8441687  
A:Accession: S30393  
A:Molecule type: mRNA  
A:Residues: 1-152 <CHA>  
A:Cross-references: UNIPROT:P25232; UNIPARC:UPI0000040B5; EMBL:X69150; NID:g38422; PIDN:  
R:Vladimirov, S.N.; Ivanov, A.V.; Karpova, G.G.; Musolyamov, A.K.; Egorov, T.A.; Thiede,  
Eur. J. Biochem. 239, 144-149, 1996  
A:Title: Characterization of the human small-ribosomal-subunit proteins by N-terminal an  
A:Reference number: S68911; MUID:96305378; PMID:8706699  
A:Accession: S68935  
A:Molecule type: protein  
A:Residues: 'S', 56-57, 'X', 59-64, 'XX', 67-69 <VLA>  
A:Cross-references: UNIPARC:UPI0000173A33  
C:Genetics:  
A:Gene: GDB:RPS18  
A:Cross-references: GDB:138447; OMIM:180473  
A:Map position: 6p21.3-6p21.3  
C:Superfamily: ribosomal protein S13/S18  
C:Keywords: blocked amino end; protein biosynthesis; ribosome  
  
Query Match 100.0%; Score 27; DB 1; Length 152;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GRGR 5  
Db 140 GRGR 144  
|||||  
  
RESULT 52  
T28833  
hypothetical protein F37C12.9 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T28833  
R:Fulton, L.  
submitted to the EMBL Data Library, March 1994  
A:Description: The sequence of C. elegans cosmid F37C12.  
A:Reference number: Z20530  
A:Accession: T28833  
A>Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-152 <FUL>  
A:Cross-references: UNIPROT:P48150; UNIPARC:UPI0000134CC5; EMBL:U000033; PIDN:AAC48301.1;  
A:Experimental source: strain Bristol N2; clone F37C12  
C:Genetics:  
A:Gene: CESP:F37C12.9  
A:Map position: 3  
A:Introns: 50/1; 88/3  
C:Superfamily: ribosomal protein S11/S14  
  
Query Match 100.0%; Score 27; DB 2; Length 152;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
|||||

Db 146 GRGR 150

RESULT 53

I76666  
ribosomal protein S18 [similarity] - mouse  
N:Alternate names: ribosomal protein S13 [mismomer]  
C:Species: Mus musculus (house mouse)  
C>Date: 02-Aug-1996 #sequence\_revision 02-Aug-1996 #text\_change 09-Jul-2004  
C:Accession: I76666  
R:MacMurray, A.J.; Shin, H.S.  
Mamm. Genome 2, 87-95, 1992  
A:Title: The murine MHC encodes a mammalian homolog of bacterial ribosomal protein S13.  
A:Reference number: I57006; MUID:92182530; PMID:1543907  
A:Accession: I76666  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-152 <RES>  
A:Cross-references: UNIPROT:P25232; UNIPARC:UPI0000040B5; GB:M76763; NID:g198579; PIDN:  
A:Accession: I57006  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-103, 'H', 105-152 <RES>  
A:Cross-references: UNIPARC:UPI000016CE74; GB:M76762; NID:g198577; PIDN:AAA16796.1; PID:  
C:Genetics:  
A:Gene: Ke-3  
A:Introns: 1/3; 34/3; 97/3; 128/2  
C:Superfamily: ribosomal protein S13/S18

Query Match 100.0%; Score 27; DB 2; Length 152;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
|||||

Db 140 GRGR 144

RESULT 54

T36954  
hypothetical protein SCJ1.21 - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: T36954  
R:Seeger, K.J.; Harris, D.; Thomson, N.R.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, September 1999  
A:Reference number: Z21607  
A:Accession: T36954  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-152 <SEE>  
A:Cross-references: UNIPROT:Q9RI23; UNIPARC:UPI00000DB393; EMBL:AL109962; PIDN:CAB53139.  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SCJ1.21  
C:Superfamily: Escherichia coli ybdQ protein

Query Match 100.0%; Score 27; DB 2; Length 152;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
|||||

Db 112 GRGR 116

RESULT 55

A56064  
ribosomal protein S14 - Chlamydomonas reinhardtii

C:Species: Chlamydomonas reinhardtii  
C>Date: 28-Apr-1995 #sequence\_revision 28-Apr-1995 #text\_change 09-Jul-2004  
C:Accession: A56064  
R:Nelson, J.A.E.; Savereide, P.B.; Lefebvre, P.A.  
Mol. Cell. Biol. 14, 4011-4019, 1994  
A:Title: The CRY1 gene in Chlamydomonas reinhardtii: structure and use as a dominant select  
A:Reference number: A56064; MUID:94254858; PMID:8196640  
A:Accession: A56064  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-153 <NEL>  
A:Cross-references: UNIPROT:P46295; UNIPARC:UPI0000134CCA; GB:U06937; NID:g463856; PIDN:  
C:Genetics:  
A:Gene: CRY1  
A:Introns: 88/3; 132/1  
C:Superfamily: ribosomal protein S11/S14

Query Match 100.0%; Score 27; DB 2; Length 153;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
|||||

Db 147 GRGR 151

RESULT 56

C70958  
hypothetical protein Rv1374c - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 09-Jul-2004  
C:Accession: C70958  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;  
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: C70958  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-153 <COL>  
A:Cross-references: UNIPROT:P71802; UNIPARC:UPI00000C1564; GB:Z81011; GB:AL123456; NID:g:  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: Rv1374c

Query Match 100.0%; Score 27; DB 2; Length 153;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
|||||

Db 37 GRGR 41

RESULT 57

G90133  
40S ribosomal protein S14 [imported] - Guillardia theta nucleomorph  
C:Species: nucleomorph Guillardia theta  
A:Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont  
C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C:Accession: G90133  
R:Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Reit  
Nature 410, 1091-1096, 2001  
A:Title: The highly reduced genome of an enslaved algal nucleus.  
A:Reference number: A99082; MUID:11323671; PMID:11323671  
A:Accession: G90133  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-159 <DOU>  
A:Cross-references: UNIPROT:Q98S22; UNIPARC:UPI00000952F6; GB:AF083031; NID:g13794381; P:

```
C:Genetics:
A:Gene: rps14
A:Map position: 3
A:Genome: nucleomorph
C:Superfamily: ribosomal protein S11/S14
C:Keywords: nucleomorph

Query Match      100.0%; Score 27; DB 2; Length 159;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5
Db 153 GRRGR 157

RESULT 58
AF3412
hypothetical protein DR6 - human herpesvirus 7 (strain JI)
C:Species: human herpesvirus 7
A:Variety: strain JI
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T41902; T42001
R:Nicholas, J.
submitted to the EMBL Data Library, December 1995
A:Description: Determination and analysis of the complete nucleotide sequence of human h
A:Reference number: Z22022
A:Accession: T41902
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-162 <N1C>
A:Cross-references: UNIPROT:Q89900; UNIPARC:UPI00000EDEC7; EMBL:U43400; PIDN:AAAC54662.1
A:Genetics: GN1
A:Accession: T42001
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-162 <N12>
A:Cross-references: UNIPARC:UPI00000EDEC7; EMBL:U43400; PIDN:AAAC54761.1
A:Genetics: GN2
C:Genetics: <GN1>
A:Gene: DR6
A:Map position: 2562-3050
C:Genetics: <GN2>
A:Gene: DR6'
A:Map position: 141609-142097

Query Match      100.0%; Score 27; DB 2; Length 162;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5
Db 44 GRRGR 48

RESULT 59
T28012
hypothetical protein ZK813.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T28012
R:Leimbach, D.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid ZK813.
A:Reference number: Z20455
A:Accession: T28012
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-163 <LEI>
A:Cross-references: UNIPROT:Q23606; UNIPARC:UPI000007ED18; EMBL:U40954; PIDN:AAAB52654.1;
A:Experimental source: strain Bristol N2; clone ZK813
C:Genetics:
A:Gene: CESP:ZK813.1
```

```
A:Map position: X
A:Introns: 36/2; 145/2

Query Match      100.0%; Score 27; DB 2; Length 163;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5
Db 54 GRRGR 58

RESULT 60
AF3412
hypothetical protein-tyrosine phosphatase BMEI1284 [imported] - Brucella melitensis (stra
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 09-Jul-2004
C:Accession: AF3412
R:DeIvecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, I.
; Mazur, M.; Goltsman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Letessc
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD3252; PMID:11756688
A:Accession: AF3412
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-169 <KUR>
A:Cross-references: UNIPROT:Q8YG77; UNIPARC:UPI0000057FED; GB:AE008917; I
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI1284
A:Map position: 1

Query Match      100.0%; Score 27; DB 2; Length 169;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5
Db 139 GRRGR 143

RESULT 61
AG1946
hypothetical protein all1122 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
C:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C:Accession: AG1946
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anat
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AG1946
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-170 <KUR>
A:Cross-references: UNIPROT:Q8YXT9; UNIPARC:UPI00000CDEB4; GB:BA000019; PIDN:BA073079.1;
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all1122

Query Match      100.0%; Score 27; DB 2; Length 170;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5
Db 126 GRRGR 130

RESULT 62
```

F87649  
ExbB/TolR family protein [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C>Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 09-Jul-2004  
C:Accession: F87649  
R:NIERMAN, W.C.; FELDBLYUM, T.V.; PAULSEN, I.T.; NELSON, K.E.; EISEN, J.; HEIDELBERG, J.  
B.; LAUB, M.T.; DEBOY, R.T.; DODSON, R.J.; DURKIN, A.S.; GWINN, M.L.; HAFT, D.H.; KOLON  
N, J.; ERMOLAeva, M.; WHITE, O.; SALZBERG, S.L.; SHAPIRO, L.; VENTER, J.C.; FRASER, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A>Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: F87649  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-172 <STO>  
A:CROSS-references: UNIPROT:Q9A3H2; UNIPARC:UPI000000C79D9; GB:AE005673; NID:g13424916; F  
C:Genetics:  
A:Gene: CC3232

Query Match 100.0%; Score 27; DB 2; Length 172;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
DB 13 GRGR 17

## RESULT 63

H87383  
hypothetical protein CC1084 [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C>Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 09-Jul-2004  
C:Accession: H87383  
R:NIERMAN, W.C.; FELDBLYUM, T.V.; PAULSEN, I.T.; NELSON, K.E.; EISEN, J.; HEIDELBERG, J.  
B.; LAUB, M.T.; DEBOY, R.T.; DODSON, R.J.; DURKIN, A.S.; GWINN, M.L.; HAFT, D.H.; KOLON  
N, J.; ERMOLAeva, M.; WHITE, O.; SALZBERG, S.L.; SHAPIRO, L.; VENTER, J.C.; FRASER, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A>Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: H87383  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-182 <STO>  
A:CROSS-references: UNIPROT:Q9A9B0; UNIPARC:UPI000000C7277; GB:AE005673; NID:g13422386; F  
C:Genetics:  
A:Gene: CC1084

Query Match 100.0%; Score 27; DB 2; Length 182;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
DB 62 GRGR 66

## RESULT 64

H87358  
hypothetical protein CC0879 [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C>Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 09-Jul-2004  
C:Accession: D87358  
R:NIERMAN, W.C.; FELDBLYUM, T.V.; PAULSEN, I.T.; NELSON, K.E.; EISEN, J.; HEIDELBERG, J.  
B.; LAUB, M.T.; DEBOY, R.T.; DODSON, R.J.; DURKIN, A.S.; GWINN, M.L.; HAFT, D.H.; KOLON  
N, J.; ERMOLAeva, M.; WHITE, O.; SALZBERG, S.L.; SHAPIRO, L.; VENTER, J.C.; FRASER, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A>Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: D87358  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-183 <STO>

A:CROSS-references: UNIPROT:Q9A9T3; UNIPARC:UPI000000C71D7; GB:AE005673; NID:g13422140; F  
C:Genetics:  
A:Gene: CC0879

Query Match 100.0%; Score 27; DB 2; Length 183;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
DB 57 GRGR 61

## RESULT 65

TFP85  
transforming protein homolog ras-85D - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C>Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 09-Jul-2004  
C:Accession: A29048  
R:BROCK, H.W.  
Gene 51, 129-137, 1987  
A>Title: Sequence and genomic structure of ras homologues Dmr85D and Dmr864B of Drosophila  
A:Reference number: A29048; MUID:87248071; PMID:3110012  
A:Accession: A29048  
A:Molecule type: DNA  
A:Residues: 1-189 <BRO>  
A:CROSS-references: UNIPROT:P08646; UNIPARC:UPI00000002D0; GB:M16429; NID:g158203; PIDN:  
A>Note: the author translated the codon GCA for residue 155 as Pro  
C:Genetics:  
A:Gene: ras-85D  
A:CROSS-references: FlyBase:FBgn0003205  
A:Introns: 47/1; 132/3  
C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
C:Keywords: GTP binding; lipoprotein; membrane protein; methylated carboxyl end; nucleot  
F;4-119/Domain: translation elongation factor Tu homology <ETU>  
F;10-17/Region: nucleotide-binding motif A (P-loop)  
F;116-119/Region: GTP-binding NKXD motif  
F;145-147/Region: GTP-binding SAK/L motif  
F;16,17,35,116,117,119,145/Binding site: Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser) #stat  
F;186/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted  
F;186/Modified site: methyl ester carboxyl end (Cys) (in mature form) #status predicted

Query Match 100.0%; Score 27; DB 1; Length 189;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5  
DB 171 GRGR 175

## RESULT 66

S35097  
transforming protein (D-ras-1) - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C>Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 09-Jul-2004  
C:Accession: S35097  
R:Neuman-Silberberg, F.S.; Schejter, E.; Hoffmann, F.M.; Shilo, B.Z.  
Cell 37, 1027-1033, 1984  
A>Title: The Drosophila ras oncogenes: structure and nucleotide sequence.  
A:Reference number: S09554; MUID:84259319; PMID:6430564  
A:Accession: S35097  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-189 <NEU>  
A:CROSS-references: UNIPROT:P08646; UNIPARC:UPI000016BD18; EMBL:K01960; NID:g158199; PIDN:  
C:Genetics:  
A:Gene: FlyBase:Ras85D  
A:CROSS-references: FlyBase:FBgn0003205  
C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
C:Keywords: GTP binding; nucleotide binding; P-loop  
F;4-119/Domain: translation elongation factor Tu homology <ETU>  
F;10-17/Region: nucleotide-binding motif A (P-loop)

F:116-119/Region: GTP-binding NKXD motif  
 F:145-147/Region: GTP-binding SAK/L motif  
 F:16,17,35,116,117,119,145/Binding site: Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser) #sta

Qy 1 GRRGR 5  
pB 171 GRRGR 175

RESULT 67

C87660

hypothetical protein CC3317 [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 09-Jul-2004

C:Accession: C87660

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolonitskii, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: C87660

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-190 <STO>

A:Cross-references: UNIPROT:Q9A387; UNIPARC:UPI00000C7A27; GB:AE005673; NID:g13425013; E

C:Genetics:

A:Gene: CC3317

Query Match	100.0%	Score 27;	DB 2;	Length 190;
Best Local Similarity	100.0%;	Pred. No. 1.9e+02;		
Matches	5;	Conservative	0;	Mismatches 0;
				Indels 0;
				Gaps 0;

Qy 1 GRRGR 5  
D<sub>b</sub> 106 GRRGR 110

RESULT 68  
B81374  
probable transposase for IS1016 NMA0554 [imported] - Neisseria meningitidis (strain Z2491  
C:Species: Neisseria meningitidis  
C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 09-Jul-2004  
C:Accession: B81974  
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel  
; Holroyd, S.; Jørgels, K.; Leachner, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
Nature 404, 502-506, 2000  
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
A:Reference number: B81775; MUID:20222556; PMID:10761919  
A:Accession: B81974  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-197 <PAR>  
A:Cross-references: UNIPROT:Q9JW43; UNIPARC:UPI000000C49C9; GB:AL157959; NID  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: NMA0554

Query Match	100.0%	Score 27;	DB 2;	Length 197;
Best Local Similarity	100.0%;	Pred. No. 1.9e+02;		
Matches	5;	Conservative	0;	Mismatches 0;
		Indels	0;	Gaps 0;

Qy 1 GRRGR 5  
80 GRRGR 84  
Dh

RESULT 69  
A81840

probable transposase for IS1016 NMA1487 [imported] - *Neisseria meningitidis* (strain Z2491)  
C:Species: *Neisseria meningitidis*  
C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 09-Jul-2004  
C:Accession: A81840  
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morelli,  
; Holroyd, S.; Jogle, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
Nature 404, 502-506, 2000  
A:Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491.  
A:Reference number: A81775; MUID:20222556; PMID:10761919  
A:Accession: A81840  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-197 <PAR>  
A:Cross-references: UNIPROT:Q9JU54; UNIPARC:UPI0000C4B88; GB:AL162756; GB:AL157959; NID:  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: NMA1487

Query Match	100.0%	Score 27;	DB 2;	Length 197;
Best Local Similarity	100.0%;	Pred. No. 1.9e+02;		
Matches	5;	Conservative	0;	Mismatches 0;
				Indels 0;
				Gaps 0;

Qy 1 GRRGR 5  
|||  
Db 80 GRRGR 84

RESULT 70

AE2808

succinoglycan biosynthesis protein exoI [imported] - Agrobacterium tumefaciens (strain C)  
 C/Species: Agrobacterium tumefaciens  
 C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 09-Jul-2004  
 C/Accession: AF2808  
 R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.;  
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan,  
 ; Karp, P.; Romero, P.; Zhang, S.  
 Science 294, 2317-2323, 2001  
 A/Authors: Yoo, H.; Tao, Y.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E.  
 aster. F.W.

A; Title: The Genome of

A;Reference number: AB2577; MUID:Z1608550; PMID:11743193  
Accession: AF2808

A;ACCESSION: AE2808  
A:STATUS: preliminary

A: Molecule type: DNA

A:Residues: 1-197 <KIR>

A:Cross-references: UNIPROT:08UE74: UNIPAR

A: Experimental source: strain C58 (Dupont)

C:Genetics:

A; Gene: exoI

A;Map position: circular chromosome

Query Match	100.0%;	Score 27;	DB 2;	Length 197;
Best Local Similarity	100.0%;	Pred. No. 1.9e+02;		

Q7 1 GRGR 5  

1  
2  
3  
4  
5  
6

RESULT 71

G72489

hypothetical protein APE2558 - Aeropyrum pernix (strain K1)

C;Species: Aeropyrum pernix

C; Date: 20-Aug-1999  
C: Accession: C73480

C/Accession: G7495  
R; Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kawauchi, T. 1999  
Astr. Res. 6: 83-101.

A; Title: Complete genome sequence of an aerobically hyper-tolerant *Escherichia coli* strain; Reference number: 373450; MIMD: 88310328; PMID: 10282866

A;Reference number: A  
A:Accession: C73488

A;ACCESSION: G/2489  
A;STATUS: preliminary

A:Molecule type: DNA  
A:Residues: 1-198 <RAW>  
A:Cross-references: UNIPROT:Q9Y885; UNIPARC:UPI000005E3C3; DDBJ:AP000064; NID:G5105945;  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE2558  
C:Superfamily: Aeropyrum pernix hypothetical protein APE2558

Query Match 100.0%; Score 27; DB 2; Length 198;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
DB 56 GRRGR 60

RESULT 72  
B81846  
Insertion element IS1016 transposase NMA1543 [imported] - Neisseria meningitidis (strain C:Species: Neisseria meningitidis  
C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 09-Jul-2004  
C:Accession: B81846  
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel  
; Holroyd, S.; Jagers, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
Nature 404, 502-506, 2000  
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
A:Reference number: A81775; MUID:20222556; PMID:10761919  
A:Accession: B81846  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-200 <PAR>  
A:Cross-references: UNIPROT:Q9JU18; UNIPARC:UPI00000C4BFF; GB:AL162756; GB:AL157959; NID  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: NMA1543  
C:Superfamily: IS1016C2 transposase NMB0583

Query Match 100.0%; Score 27; DB 2; Length 200;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
DB 85 GRRGR 89

RESULT 73  
AH0120  
probable membrane protein YPO0983 [imported] - Yersinia pestis (strain CO92)  
C:Species: Yersinia pestis  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Jul-2004  
C:Accession: AH0120  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,  
Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AH0120  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-216 <KUR>  
A:Cross-references: UNIPROT:Q8ZHB8; UNIPARC:UPI00000DCE72; GB:AL590842; PIDN:CAC89827.1;  
C:Genetics:  
A:Gene: YPO0983  
C:Superfamily: Mycobacterium tuberculosis hypothetical protein Rv0658c

Query Match 100.0%; Score 27; DB 2; Length 216;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5

DB 146 GRRGR 150

RESULT 74  
T47329  
hypothetical protein T12K4.160 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 09-Jul-2004  
C:Accession: T47329  
R:Monfort, A.; Casacuberta, E.; Puigdomenech, P.; Mewes, H.W.; Rudd, S.; Lemcke, K.; Maye;  
submitted to the Protein Sequence Database, April 2000  
A:Reference number: Z2460  
A:Accession: T47329  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-216 <MON>  
A:Cross-references: UNIPROT:Q9M293; UNIPARC:UPI000009FF1B; EMBL:AL138640  
A:Experimental source: cultivar Columbia; BAC clone T12K4  
C:Genetics:  
A:Map position: 3  
A:Introns: 106/1  
A:Note: T12K4.160  
C:Superfamily: Arabidopsis thaliana hypothetical protein T12K4.160

Query Match 100.0%; Score 27; DB 2; Length 216;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
DB 35 GRRGR 39

RESULT 75  
GB1087  
IS1016C2 transposase NMB1411 [imported] - Neisseria meningitidis (strain MC58 serogroup I  
C:Species: Neisseria meningitidis  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 09-Jul-2004  
C:Accession: GB1087  
R:Tetzelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.  
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Fizza, M.  
Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ver  
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
A:Reference number: A81000; MUID:20175755; PMID:10710307  
A:Accession: GB1087  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-217 <TET>  
A:Cross-references: UNIPROT:Q9JYV8; UNIPARC:UPI00000C46DC; GB:AE002489; GB:AE002098; NID:  
A:Experimental source: serogroup B, strain MC58  
C:Genetics:  
A:Gene: NMB1411  
C:Superfamily: IS1016C2 transposase NMB0583

Query Match 100.0%; Score 27; DB 2; Length 217;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
DB 80 GRRGR 84

Search completed: December 2, 2005, 10:07:34  
Job time : 39 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2005, 09:42:09 ; Search time 147 Seconds  
(without alignments)  
23.998 Million cell updates/sec

Title: SEQ-GRGR  
Perfect score: 27  
Sequence: 1 grgr 5

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 1000 summaries

Database : Uniprot\_05.80.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	27	100.0	17	2	Q6SV46	chlamydomon
2	27	100.0	24	1	HSP3	octopus vul
3	27	100.0	24	2	Q6TQ76	saccharomyc
4	27	100.0	28	1	HSP4	octopus vul
5	27	100.0	30	1	HSP5	octopus vul
6	27	100.0	37	1	PRZ3	scyliorhinu
7	27	100.0	45	2	Q7LZB0	chrysemys p
8	27	100.0	45	2	Q7LZB1	chrysemys p
9	27	100.0	45	2	Q7LZB5	chrysemys p
10	27	100.0	46	2	Q9GT46	anopheles g
11	27	100.0	47	2	Q7LZA7	chrysemys p
12	27	100.0	48	2	Q9GMB9	oryctolagus
13	27	100.0	48	2	Q6ZCY0	burkholderi
14	27	100.0	49	2	Q4GXQ0	eucinetus s
15	27	100.0	50	2	Q4X611	plasmodium
16	27	100.0	51	2	Q7LZL4	xenopus lae
17	27	100.0	53	2	Q8RYU2	oryza sativ
18	27	100.0	54	2	Q45362	bordetella
19	27	100.0	54	2	Q4KKI8	pseudomonas
20	27	100.0	56	1	HSP1	octopus vul
21	27	100.0	57	1	HSP1	didelphis m
22	27	100.0	57	1	HSP1	monodelphis
23	27	100.0	58	1	HSP1	chrysemys p
24	27	100.0	59	2	Q981D3	sulfolobus
25	27	100.0	60	2	Q7QZW6	giardia lam
26	27	100.0	63	2	Q4T4G0	tetradon n
27	27	100.0	64	2	Q6ZCY1	oryza sativ
28	27	100.0	64	2	Q4RQ31	tetradon n
29	27	100.0	66	1	RS14	nicotiana t
30	27	100.0	66	2	Q6TWF2	oryza sativ
31	27	100.0	68	2	Q8S5M9	oryza sativ

32	27	100.0	70	2	Q9FRQ5	oryza sativ
33	27	100.0	70	2	Q9QVVO	mus sp. hox
34	27	100.0	70	2	Q9PRL3	gallus sp.
35	27	100.0	71	2	Q4TTN3	streptomyce
36	27	100.0	73	2	Q6K3X4	oryza sativ
37	27	100.0	74	2	Q69VE1	oryza sativ
38	27	100.0	74	2	Q8SAY4	oryza sativ
39	27	100.0	75	2	Q54391	streptomyce
40	27	100.0	76	2	Q6ZFA5	oryza sativ
41	27	100.0	76	2	Q8FS22	corynebacte
42	27	100.0	77	2	Q67PE1	syntrophobacte
43	27	100.0	79	2	Q7SHB4	neurospora
44	27	100.0	79	2	Q9XSU9	canis famli
45	27	100.0	79	2	Q6ERZ5	oryza sativ
46	27	100.0	80	2	Q5YS36	nocardia fa
47	27	100.0	80	2	Q9YMW8	lymantaria d
48	27	100.0	81	2	Q73ZQ7	mycobacteri
49	27	100.0	82	1	Y696	borrelia bu
50	27	100.0	82	2	Q4VXZ3	homo sapien
51	27	100.0	82	2	Q660H5	borrelia ga
52	27	100.0	84	2	Q6H6S6	oryza sativ
53	27	100.0	85	2	Q48451	bacterioph
54	27	100.0	85	2	Q6H835	oryza sativ
55	27	100.0	85	2	Q801B9	latimeria m
56	27	100.0	85	2	Q801D1	latimeria m
57	27	100.0	85	2	Q533W6	oreochromis
58	27	100.0	86	2	Q50112	mycobacteri
59	27	100.0	86	2	Q4SQ17	tetradon n
60	27	100.0	88	2	Q9LIX6	oryza sativ
61	27	100.0	89	2	Q8FRF8	corynebacte
62	27	100.0	91	2	Q53PD9	oryza sativ
63	27	100.0	92	2	Q51CD7	entamoeba h
64	27	100.0	92	2	Q5CMB2	cryptospori
65	27	100.0	93	2	Q8WVG6	homo sapien
66	27	100.0	93	2	Q5N9T5	oryza sativ
67	27	100.0	93	2	Q4J3J0	azotobacter
68	27	100.0	94	2	Q86MX2	dictyosteli
69	27	100.0	94	2	Q5NB13	oryza sativ
70	27	100.0	94	2	Q7SUY3	human immu
71	27	100.0	94	2	Q8UMG3	human immu
72	27	100.0	95	2	Q6EQW9	oryza sativ
73	27	100.0	95	2	Q9Z291	cricetulus
74	27	100.0	95	2	Q6WS73	human immu
75	27	100.0	95	2	Q97061	human immu
76	27	100.0	95	2	Q9QML1	human immu
77	27	100.0	96	2	Q6JNE4	human immu
78	27	100.0	96	2	Q72610	human immu
79	27	100.0	96	2	Q72615	human immu
80	27	100.0	96	2	Q58Q79	human immu
81	27	100.0	96	2	Q596M9	human immu
82	27	100.0	96	2	Q596N4	human immu
83	27	100.0	96	2	Q596P0	human immu
84	27	100.0	96	2	Q596P6	human immu
85	27	100.0	96	2	Q596Q1	human immu
86	27	100.0	96	2	Q596R0	human immu
87	27	100.0	96	2	Q596R9	human immu
88	27	100.0	96	2	Q596S8	human immu
89	27	100.0	96	2	Q596T7	human immu
90	27	100.0	96	2	Q5MGU0	human immu
91	27	100.0	96	2	Q5MH16	human immu
92	27	100.0	96	2	Q5VCT1	human immu
93	27	100.0	96	2	Q5VGQ8	human immu
94	27	100.0	96	2	Q6JN99	human immu
95	27	100.0	96	2	Q6JNB7	human immu
96	27	100.0	96	2	Q6JNF3	human immu
97	27	100.0	96	2	Q6JNG2	human immu
98	27	100.0	96	2	Q6JNH1	human immu
99	27	100.0	96	2	Q6JN17	human immu
100	27	100.0	96	2	Q6JNJ6	human immu
101	27	100.0	96	2	Q6JNL4	human immu
102	27	100.0	96	2	Q6JNM2	human immu
103	27	100.0	96	2	Q6JNM8	human immu
104	27	100.0	96	2	Q6JNN7	human immu

105	27	100.0	96	2	Q6JNP6_9HIV1	Q6jnp6	human	immun	178	27	100.0	96	2	Q9DKH0_9HIV1	Q9dkh0	human	immun
106	27	100.0	96	2	Q6JNT2_9HIV1	Q6jnt2	human	immun	179	27	100.0	96	2	Q9IMJ5_9HIV1	Q9imj5	human	immun
107	27	100.0	96	2	Q6JNU1_9HIV1	Q6jnu1	human	immun	180	27	100.0	96	2	Q9IV16_9HIV1	Q9iv16	human	immun
108	27	100.0	96	2	Q6JNW8_9HIV1	Q6jnw8	human	immun	181	27	100.0	96	2	Q9IW41_9HIV1	Q9iw41	human	immun
109	27	100.0	96	2	Q6JNX7_9HIV1	Q6jnx7	human	immun	182	27	100.0	96	2	Q9IW50_9HIV1	Q9iw50	human	immun
110	27	100.0	96	2	Q6JNY6_9HIV1	Q6jny6	human	immun	183	27	100.0	96	2	Q9Q6H4_9HIV1	Q9q6h4	human	immun
111	27	100.0	96	2	Q6JP04_9HIV1	Q6jp04	human	immun	184	27	100.0	96	2	Q9Q611_9HIV1	Q9q611	human	immun
112	27	100.0	96	2	Q6JP22_9HIV1	Q6jp22	human	immun	185	27	100.0	96	2	Q9Q617_9HIV1	Q9q617	human	immun
113	27	100.0	96	2	Q6JEM6_9HIV1	Q6jem6	human	immun	186	27	100.0	96	2	Q9Q6F7_9HIV1	Q9q6f7	human	immun
114	27	100.0	96	2	Q6JER1_9HIV1	Q6jer1	human	immun	187	27	100.0	96	2	Q9QML5_9HIV1	Q9qml5	human	immun
115	27	100.0	96	2	Q6JET7_9HIV1	Q6jet7	human	immun	188	27	100.0	96	2	Q9QRX5_9HIV1	Q9qrx5	human	immun
116	27	100.0	96	2	Q6JEU6_9HIV1	Q6jeu6	human	immun	189	27	100.0	96	2	Q9W9K9_9HIV1	Q9w9k9	human	immun
117	27	100.0	96	2	Q6JF00_9HIV1	Q6jf00	human	immun	190	27	100.0	96	2	Q9WRE7_9HIV1	Q9wre7	human	immun
118	27	100.0	96	2	Q6JF08_9HIV1	Q6jf08	human	immun	191	27	100.0	96	2	Q9WRP4_9HIV1	Q9wrf4	human	immun
119	27	100.0	96	2	Q6JF15_9HIV1	Q6jf15	human	immun	192	27	100.0	96	2	Q9WRG0_9HIV1	Q9wrg0	human	immun
120	27	100.0	96	2	Q6JF24_9HIV1	Q6jf24	human	immun	193	27	100.0	96	2	Q9WSM7_9HIV1	Q9wsm7	human	immun
121	27	100.0	96	2	Q6JF33_9HIV1	Q6jf33	human	immun	194	27	100.0	96	2	Q9WSP4_9HIV1	Q9wsp4	human	immun
122	27	100.0	96	2	Q6JF42_9HIV1	Q6jf42	human	immun	195	27	100.0	96	2	Q9WSQ0_9HIV1	Q9wsq0	human	immun
123	27	100.0	96	2	Q6JF51_9HIV1	Q6jf51	human	immun	196	27	100.0	96	2	Q4QXJ3_9HIV1	Q4qxj3	human	immun
124	27	100.0	96	2	Q6JF60_9HIV1	Q6jf60	human	immun	197	27	100.0	96	2	Q4QX15_9HIV1	Q4qx15	human	immun
125	27	100.0	96	2	Q6JF67_9HIV1	Q6jf67	human	immun	198	27	100.0	96	2	Q4QX84_9HIV1	Q4qx84	human	immun
126	27	100.0	96	2	Q6JF91_9HIV1	Q6jf91	human	immun	199	27	100.0	96	2	Q4QX75_9HIV1	Q4qx75	human	immun
127	27	100.0	96	2	Q6JF9A_9HIV1	Q6jfa9	human	immun	200	27	100.0	96	2	Q4QX66_9HIV1	Q4qx66	human	immun
128	27	100.0	96	2	Q6JF8B_9HIV1	Q6jfb8	human	immun	201	27	100.0	96	2	Q4QX58_9HIV1	Q4qx58	human	immun
129	27	100.0	96	2	Q6JFDE_9HIV1	Q6jfd6	human	immun	202	27	100.0	96	2	Q4QX42_9HIV1	Q4qx42	human	immun
130	27	100.0	96	2	Q6JFEO_9HIV1	Q6jfe0	human	immun	203	27	100.0	96	2	Q4QX34_9HIV1	Q4qx34	human	immun
131	27	100.0	96	2	Q6JFF5_9HIV1	Q6jff5	human	immun	204	27	100.0	96	2	Q4QX26_9HIV1	Q4qx26	human	immun
132	27	100.0	96	2	Q6JFH3_9HIV1	Q6jfh3	human	immun	205	27	100.0	97	2	Q4QX85_CANFA	Q4x85	canis	famil
133	27	100.0	96	2	Q6JFI2_9HIV1	Q6jfi2	human	immun	206	27	100.0	97	2	Q6Z962_ORISA	Q6z962	oryza	sativ
134	27	100.0	96	2	Q6JFJ1_9HIV1	Q6jfj1	human	immun	207	27	100.0	97	2	Q5VCU4_9HIV1	Q5vcu4	human	immun
135	27	100.0	96	2	Q6JFK0_9HIV1	Q6jfk0	human	immun	208	27	100.0	98	2	Q6JNA8_9HIV1	Q6jna8	human	immun
136	27	100.0	96	2	Q6JFL7_9HIV1	Q6jfl7	human	immun	209	27	100.0	98	2	Q6JFK9_9HIV1	Q6jfk9	human	immun
137	27	100.0	96	2	Q6JFMA_9HIV1	Q6jfm4	human	immun	210	27	100.0	99	2	Q6JNC6_9HIV1	Q6jnc6	human	immun
138	27	100.0	96	2	Q6JFNE_9HIV1	Q6jfn2	human	immun	211	27	100.0	99	2	Q6JNKS_9HIV1	Q6jnk5	human	immun
139	27	100.0	96	2	Q6JFQ0_9HIV1	Q6jfq0	human	immun	212	27	100.0	99	2	Q8JBK2_9HIV1	Q8jbk2	human	immun
140	27	100.0	96	2	Q6X6N8_9HIV1	Q6x6n8	human	immun	213	27	100.0	100	2	Q7XHM9_ORISA	Q7xhm9	oryza	sativ
141	27	100.0	96	2	Q6X6V7_9HIV1	Q6x6v7	human	immun	214	27	100.0	100	2	Q4LWQ0_9BURK	Q4lwq0	burkholderi	
142	27	100.0	96	2	Q70XC8_9HIV1	Q70xc8	human	immun	215	27	100.0	101	2	Q998E9_9HIV1	Q998e9	human	immun
143	27	100.0	96	2	Q74455_9HIV1	Q74455	human	immun	216	27	100.0	101	2	Q6ZH07_ORISA	Q6zh07	oryza	sativ
144	27	100.0	96	2	Q7SKG3_9HIV1	Q7skg8	human	immun	217	27	100.0	102	2	Q8YXW8_ORISA	Q8yxw8	oryza	sativ
145	27	100.0	96	2	Q7SKH7_9HIV1	Q7skh7	human	immun	218	27	100.0	102	2	Q8S150_ORISA	Q8s150	oryza	sativ
146	27	100.0	96	2	Q7SV26_9HIV1	Q7sv26	human	immun	219	27	100.0	102	2	Q8PL81_XANAC	Q8pl81	xanthomonas	
147	27	100.0	96	2	Q7ZBG4_9HIV1	Q7zbg4	human	immun	220	27	100.0	102	2	Q7SKF9_9HIV1	Q7skf9	human	immun
148	27	100.0	96	2	Q7ZJ15_9HIV1	Q7zj15	human	immun	221	27	100.0	103	2	Q69K26_ORISA	Q69k26	oryza	sativ
149	27	100.0	96	2	Q8ADF5_9HIV1	Q8adf5	human	immun	222	27	100.0	103	2	Q6YPH1_ORISA	Q6yph1	oryza	sativ
150	27	100.0	96	2	Q8ADX3_9HIV1	Q8adx3	human	immun	223	27	100.0	103	2	Q6F9U8_9HOO	Q6f9u8	azearcus	ev
151	27	100.0	96	2	Q8AE38_9HIV1	Q8ae38	human	immun	224	27	100.0	103	2	Q6G4S9_BARHE	Q6g4s9	bartonella	
152	27	100.0	96	2	Q8AK04_9HIV1	Q8ak04	human	immun	225	27	100.0	103	2	Q99H59_BRAJA	Q99h59	bradyrhizob	
153	27	100.0	96	2	Q8AK13_9HIV1	Q8ak13	human	immun	226	27	100.0	105	2	Q6FNP3_CANGA	Q6fnp3	candida	glia
154	27	100.0	96	2	Q8J3U7_9HIV1	Q8j3u7	human	immun	227	27	100.0	105	2	Q5QMA8_ORISA	Q5qma8	oryza	sativ
155	27	100.0	96	2	Q8J9B3_9HIV1	Q8j9b3	human	immun	228	27	100.0	106	2	Q9YAL4_AERPE	Q9yal4	aeropyrum	p
156	27	100.0	96	2	Q8J9C3_9HIV1	Q8j9c2	human	immun	229	27	100.0	106	2	Q6ZDB9_ORISA	Q6zdb9	oryza	sativ
157	27	100.0	96	2	Q8JBG6_9HIV1	Q8jbg6	human	immun	230	27	100.0	107	2	Q5ZBS2_ORISA	Q5zbs2	oryza	sativ
158	27	100.0	96	2	Q8JB14_9HIV1	Q8jb14	human	immun	231	27	100.0	107	2	Q6K2Y7_ORISA	Q6k2y7	oryza	sativ
159	27	100.0	96	2	Q8JBPO_9HIV1	Q8jbpo	human	immun	232	27	100.0	108	2	Q96HT0_HUMAN	Q96ht0	homo	sapien
160	27	100.0	96	2	Q8JBQ0_9HIV1	Q8jbq0	human	immun	233	27	100.0	108	2	Q8W332_ORISA	Q8w332	oryza	sativ
161	27	100.0	96	2	Q8JBT6_9HIV1	Q8jbt6	human	immun	234	27	100.0	109	2	Q6Z8F0_ORISA	Q6z8f0	oryza	sativ
162	27	100.0	96	2	Q8JBUS_9HIV1	Q8jbus	human	immun	235	27	100.0	110	2	Q95SY6_DROME	Q95sy6	drosophila	
163	27	100.0	96	2	Q8JBV4_9HIV1	Q8jbv4	human	immun	236	27	100.0	110	2	Q8KW22_9HOB	Q8kw22	ruegeria	sp
164	27	100.0	96	2	Q8JC21_9HIV1	Q8jc21	human	immun	237	27	100.0	111	2	Q5QLU0_ORISA	Q5qlu0	oryza	sativ
165	27	100.0	96	2	Q8Q2F6_9HIV1	Q8q2f6	human	immun	238	27	100.0	111	2	Q8H4W2_ORISA	Q8h4w2	oryza	sativ
166	27	100.0	96	2	Q8QFK7_9HIV1	Q8qfk7	human	immun	239	27	100.0	111	2	Q92MB6_RHIME	Q92mb6	rhizobium	m
167	27	100.0	96	2	Q8QFK9_9HIV1	Q8qfk9	human	immun	240	27	100.0	112	2	Q4J232_AZOVI	Q4j232	azotobacter	
168	27	100.0	96	2	Q8QMH2_9HIV1	Q8qmh2	human	immun	241	27	100.0	112	2	Q86694_STRCO	Q86694	streptomyce	
169	27	100.0	96	2	Q90CK1_9HIV1	Q90ck1	human	immun	242	27	100.0	112	2	Q4S6M5_TETNG	Q4s6m5	tetradodon	n
170	27	100.0	96	2	Q90CLO_9HIV1	Q90clo	human	immun	243	27	100.0	113	2	Q9UJ34_HUMAN	Q9uj34	homo	sapien
171	27	100.0	96	2	Q90D19_9HIV1	Q90d19	human	immun	244	27	100.0	113	2	Q6H587_ORISA	Q6h587	oryza	sativ
172	27	100.0	96	2	Q90DY9_9HIV1	Q90dy9	human	immun	245	27	100.0	113	2	Q9XCB3_RHOMR	Q9xcb3	rhodothermu	
173	27	100.0	96	2	Q90DZ9_9HIV1	Q90dz9	human	immun	246	27	100.0	114	2	Q60MS0_CAEBR	Q60ms0	caenorhabdi	
174	27	100.0	96	2	Q90VT3_9HIV1	Q90vt2	human	immun	247	27	100.0	114	2	Q6Z134_ORISA	Q6z134	oryza	sativ
175	27	100.0	96	2	Q998H5_9HIV1	Q998h5	human	immun	248	27	100.0	114	2	Q5DLU5_9BACT	Q5dlu5	uncultured	
176	27	100.0	96	2	Q9DH18_9HIV1	Q9dh18	human	immun	249	27	100.0	116	2	Q7R0E1_GALIA	Q7r0e1	giardia	lam
177	27	100.0	96	2	Q9DKD7_9HIV1	Q9dkd7	human	immun	250	27	100.0	116	2	Q5MGK3_9NEOP	Q5mgk3	lonomia	obi

251	27	100.0	116	2	Q6ZF07_ORYSA	Q6zf07 oryza sativ	324	27	100.0	137	1	RS11_PYRFU	Q8ue03 pyrococcus
252	27	100.0	116	2	Q53LB5_ORYSA	Q53lb5 oryza sativ	325	27	100.0	137	1	RS11_PYRHO	P62011 pyrococcus
253	27	100.0	117	2	Q8TW27_METRA	Q8tw27 methanopyru	326	27	100.0	137	1	RS14B_YEAST	P39516 saccharomyc
254	27	100.0	117	2	Q84SP6_ORYSA	Q84sp6 oryza sativ	327	27	100.0	137	1	RS14_KLULA	P27069 kluyveromyc
255	27	100.0	117	2	Q9K2L3_STRCO	Q9k2l3 streptomyc	328	27	100.0	137	1	Q69V52_ORYSA	Q69v52 oryza sativ
256	27	100.0	117	2	Q4PR95_BRARE	Q4pr95 brachydanio	329	27	100.0	138	2	Q6FNE1_CANGA	Q6fne1 candida gla
257	27	100.0	118	2	Q5D3Q8_9BACT	Q5d3q8 uncultured	330	27	100.0	138	2	Q758E0_ASHGO	Q758e0 ashbya gos
258	27	100.0	118	2	Q5D3R9_9BACT	Q5d3r9 uncultured	331	27	100.0	138	2	Q8N8G7_HUMAN	Q8n8g7 homo sapien
259	27	100.0	118	2	Q5F3V7_CHICK	Q5f3v7 gallus gall	332	27	100.0	138	2	Q5Z8S8_ORYSA	Q5z8s8 oryza sativ
260	27	100.0	120	2	Q71D43_DROYA	Q71d43 drosophila	333	27	100.0	138	2	Q7X8A7_ORYSA	Q7x8a7 oryza sativ
261	27	100.0	120	2	Q6K5Z0_ORYSA	Q6k5z0 oryza sativ	334	27	100.0	139	1	RS14_SCHPO	O14150 schizosacch
262	27	100.0	120	2	Q7XW02_ORYSA	Q7xw02 oryza sativ	335	27	100.0	139	1	Q6BHV6_DEBHA	Q6bhv6 debaryomyc
263	27	100.0	120	2	Q5YET5_CHLS6	Q5yet5 chlorarachn	336	27	100.0	139	2	Q26948_TRYCR	Q26948 trypanosoma
264	27	100.0	121	2	Q8N9K7_HUMAN	Q8n9k7 homo sapien	337	27	100.0	139	2	Q6L4Q3_ORYSA	Q6l4q3 oryza sativ
265	27	100.0	121	2	Q67UR0_ORYSA	Q67ur0 oryza sativ	338	27	100.0	140	1	RS11_PYRKO	O5j5f3 pyrococcus
266	27	100.0	121	2	Q69SM6_ORYSA	Q69sm6 oryza sativ	339	27	100.0	140	2	Q84JE6_ORYSA	Q84je6 oryza sativ
267	27	100.0	121	2	Q6KAI3_ORYSA	Q6kai3 oryza sativ	340	27	100.0	140	2	Q6Y200_PAGMA	Q6y200 pagrus majo
268	27	100.0	121	2	Q7F8R7_ORYSA	Q7f8r7 oryza sativ	341	27	100.0	142	2	Q9GPA9_CHEDE	Q9gpa9 cherax deat
269	27	100.0	121	2	Q93A26_9PROT	Q93a26 pseudomonas	342	27	100.0	142	2	Q6ZBK8_ORYSA	Q6zbk8 oryza sativ
270	27	100.0	122	2	Q6ATQ4_ORYSA	Q6atq4 oryza sativ	343	27	100.0	143	2	Q7PRN3_ANOGA	Q7prn3 anopheles g
271	27	100.0	122	2	Q6EQ47_ORYSA	Q6eq47 oryza sativ	344	27	100.0	143	2	Q516B8_ENTHI	Q516b8 entamoeba h
272	27	100.0	122	2	P72390_STRCO	P72390 streptomyc	345	27	100.0	143	2	Q8LGM3_ORYSA	Q8lgm3 oryza sativ
273	27	100.0	122	2	Q49710_MYCLE	Q49710 mycobacteri	346	27	100.0	143	2	Q53K00_ORYSA	Q53k00 oryza sativ
274	27	100.0	124	2	Q6Z1E4_ORYSA	Q6z1e4 oryza sativ	347	27	100.0	143	2	Q9CJW0_MOUSE	Q9cjw0 mus musculu
275	27	100.0	124	2	Q9FRG1_ORYSA	Q9frg1 oryza sativ	348	27	100.0	144	2	Q6Y776_ORYSA	Q6y776 oryza sativ
276	27	100.0	125	2	Q8NEE2_HUMAN	Q8nee2 homo sapien	349	27	100.0	144	2	Q5VR13_ORYSA	Q5vr13 oryza sativ
277	27	100.0	125	2	Q7PUG1_ANOGA	Q7pug1 anopheles g	350	27	100.0	145	2	Q76KS3_GIALA	Q76ks3 giardia lam
278	27	100.0	125	2	Q655T4_ORYSA	Q655t4 oryza sativ	351	27	100.0	145	2	Q7QSF3_GIALA	Q7qsf3 giardia lam
279	27	100.0	125	2	Q56XY0_ARATH	Q56xy0 arabidopsis	352	27	100.0	145	2	Q21606_CAEEL	Q21606 caenorhabdi
280	27	100.0	125	2	Q6ZLN3_ORYSA	Q6zln3 oryza sativ	353	27	100.0	145	2	Q5N9Q1_ORYSA	Q5n9q1 oryza sativ
281	27	100.0	126	2	Q5DD12_SCHUA	Q5dd12 schistosoma	354	27	100.0	145	2	Q5Z6A8_ORYSA	Q5z6a8 oryza sativ
282	27	100.0	126	2	Q5ZB28_ORYSA	Q5zb28 oryza sativ	355	27	100.0	145	2	Q6YZ24_ORYSA	Q6yz24 oryza sativ
283	27	100.0	126	2	Q8S3P6_ORYSA	Q8s3p6 oryza sativ	356	27	100.0	145	2	Q7XKW8_ORYSA	Q7xkw8 oryza sativ
284	27	100.0	127	1	RS11_PTCO	Q6kzp6 picophilus	357	27	100.0	146	2	Q513Z8_ENTHI	Q513z8 entamoeba h
285	27	100.0	127	2	Q4WLH1_ASPEU	Q4wlh1 aspergillus	358	27	100.0	146	2	Q6EP47_ORYSA	Q6ep47 oryza sativ
286	27	100.0	127	2	Q6H4P1_ORYSA	Q6h4p1 oryza sativ	359	27	100.0	146	2	Q7XJU0_SECE	Q7xju0 secale cere
287	27	100.0	127	2	Q6K4Z9_ORYSA	Q6k4z9 oryza sativ	360	27	100.0	147	2	Q6XMY4_RHOER	Q6xmy4 rhodococcc
288	27	100.0	127	2	Q6JAH3_SORBI	Q6jah3 sorghum bic	361	27	100.0	148	2	Q82D22_STRAW	Q82d22 streptomyc
289	27	100.0	128	2	Q5QN79_ORYSA	Q5qn79 oryza sativ	362	27	100.0	149	1	RS141_MAIZE	P19550 zea mays (m
290	27	100.0	128	2	Q5P6N6_AZOSE	Q5p6n6 azoarcus sp	363	27	100.0	149	2	Q5B0H0_EMENI	Q5b0h0 aspergillus
291	27	100.0	128	2	Q825S8_STRAM	Q825s8 streptomyc	364	27	100.0	149	2	Q5DG39_SCHJA	Q5dg39 schistosoma
292	27	100.0	129	2	Q9NT41_HUMAN	Q9nt41 homo sapien	365	27	100.0	149	2	Q6ESN4_ORYSA	Q6esn4 oryza sativ
293	27	100.0	129	2	Q8LN23_ORYSA	Q8ln23 oryza sativ	366	27	100.0	149	2	Q6H706_ORYSA	Q6h706 oryza sativ
294	27	100.0	129	2	Q9K125_NEIMB	Q9k125 neisseria m	367	27	100.0	149	2	Q6L526_ORYSA	Q6l526 oryza sativ
295	27	100.0	130	2	Q8LH69_ORYSA	Q8lh69 oryza sativ	368	27	100.0	149	2	Q6ZAM9_ORYSA	Q6zam9 oryza sativ
296	27	100.0	131	1	RS11_ASPE	Q9yb55 aeropyrum p	369	27	100.0	149	2	Q7XU68_ORYSA	Q7xu68 oryza sativ
297	27	100.0	131	2	Q54V32_DICDI	Q54v32 dictyosteli	370	27	100.0	150	1	RS141_ARATH	Q9sih0 arabidopsis
298	27	100.0	131	2	Q5C222_CRYPV	Q5c222 cryptospori	371	27	100.0	150	1	RS142_ARATH	Q9cax6 arabidopsis
299	27	100.0	131	2	Q9SM13_MAIZE	Q9sm13 zea mays (m	372	27	100.0	150	1	RS142_MAIZE	P19551 zea mays (m
300	27	100.0	132	1	RS11_SULAC	P39469 sulfolobus	373	27	100.0	150	1	RS143_ARATH	P42036 arabidopsis
301	27	100.0	132	1	RS11_SULTO	P95988 sulfolobus	374	27	100.0	150	1	RS14_CRIGR	P62265 cricetulus
302	27	100.0	132	1	RS11_SULTO	Q96yv9 oryza sativ	375	27	100.0	150	1	RS14_HUMAN	P62263 homo sapien
303	27	100.0	132	2	Q52709_ORYSA	Q52709 oryza sativ	376	27	100.0	150	1	RS14_LUPLU	Q22584 lupinus lut
304	27	100.0	132	2	Q6EPP4_ORYSA	Q6ep4 oryza sativ	377	27	100.0	150	1	RS14_MOUSE	P62264 mus musculu
305	27	100.0	132	2	Q7NXX3_CHRVO	Q7nxx3 chromobacte	378	27	100.0	150	1	RS14_NEUCR	P19115 neurospora
306	27	100.0	132	2	Q4TAG6_TETNG	Q4tag6 tetraodon n	379	27	100.0	150	2	Q55QD7_CRYNE	P13471 rattus norv
307	27	100.0	133	2	Q69XD2_ORYSA	Q69xd2 oryza sativ	380	27	100.0	150	2	Q5EN09_MAGGR	Q55qd7 cryptococc
308	27	100.0	133	2	Q5ZCM6_ORYSA	Q5zcm6 oryza sativ	381	27	100.0	150	2	Q520W2_MAGGR	Q5en09 magnaporth
309	27	100.0	134	1	RS14_CANAL	Q96w53 candida alb	382	27	100.0	150	2	Q5KFT0_CRYNE	Q520w2 magnaporth
310	27	100.0	134	1	RS14_TORRU	Q9xek6 tortula rur	383	27	100.0	150	2	Q4XIC6_ASPFU	Q5kft0 cryptococc
311	27	100.0	134	2	Q94H27_ORYSA	Q94h27 oryza sativ	384	27	100.0	150	2	Q4KTC8_SUBDO	Q4xic6 aspergillus
312	27	100.0	134	2	Q8VK63_MYCTU	Q8vk63 mycobacteri	385	27	100.0	150	2	Q4YXJ8_PLABE	Q4kct8 suberites d
313	27	100.0	134	2	Q5WY76_LEGPL	Q5wy76 legionella	386	27	100.0	150	2	Q4YXJ8_PLABE	Q4yxj8 plasmodium
314	27	100.0	134	2	Q5X7E0_LEGPA	Q5x7e0 legionella	387	27	100.0	150	2	Q6H7T1_ORYSA	Q6h7t1 oryza sativ
315	27	100.0	134	2	Q5ZAW7_LEGPH	Q5zaw7 legionella	388	27	100.0	150	2	Q84X98_BRANA	Q84x98 brassica na
316	27	100.0	134	2	Q7TN26_MOUSE	Q7tn26 mus musculu	389	27	100.0	150	2	Q75RT9_9BACT	Q75rt9 uncultured
317	27	100.0	135	2	Q6FKW5_CANGA	Q6fkw5 candida gla	390	27	100.0	150	2	Q8VAA8_WSSV	Q8vaa8 white spot
318	27	100.0	135	2	Q4H4C2_9PROT	Q4h4c2 azospirillum	391	27	100.0	151	2	RS14_DROME	P14130 drosophila
319	27	100.0	136	1	RS14A_YEAST	P06367 saccharomyc	392	27	100.0	151	1	RS14_PODCA	Q08699 podocoryne
320	27	100.0	136	2	Q75KY9_ORYSA	Q75ky9 oryza sativ	393	27	100.0	151	1	RS14_PROCL	P48955 procamburus
321	27	100.0	136	2	Q9D717_MOUSE	Q9d717 mus musculu	394	27	100.0	151	2	Q41J77_GIBZE	Q41j77 gibberella
322	27	100.0	137	1	PA23_TRIOR	Q92152 trimeresuru	395	27	100.0	151	2	Q5BU10_HUMAN	Q5bu10 homo sapien
323	27	100.0	137	1	RS11_PYRAB	P62010 pyrococcus	396	27	100.0	151	2	Q61159_DROME	Q61159 drosophila

397	27	100.0	151	2	Q56FC4_9HYME	Q56fc4_lysiphebus	470	27	100.0	152	2	Q5DVH4_PLAFA	Q5dvh4_platichthys
398	27	100.0	151	2	Q56J82_9BILA	Q56j82_philodina s	471	27	100.0	152	2	Q6NTT2_XENLA	Q6ntt2_xenopus lae
399	27	100.0	151	2	Q5UAM9_BOMMO	Q5uam9_bombyx mori	472	27	100.0	153	1	RS14_CHLSE	Q64295 chlamydomon
400	27	100.0	151	2	Q66SW2_BOMMO	Q66sw2_bombyx mori	473	27	100.0	153	1	RS18_ENCCU	Q8srp2 encephalico
401	27	100.0	151	2	Q6B8A8_9ACAR	Q6b8a8_ixodes paci	474	27	100.0	153	2	Q4PM79_IXOSC	Q4pm79_ixodes scap
402	27	100.0	151	2	Q6E8UY_9COLE	Q6e8uy_dascillus c	475	27	100.0	153	2	Q654W5_ORYSA	Q654w5_oryza sativ
403	27	100.0	151	2	Q6F434_PLUXY	Q6f434_plutella xy	476	27	100.0	154	2	Q4PFQ3_USTMA	Q4pfq3_ustilago ma
404	27	100.0	151	2	Q6IV88_BRABE	Q6iv88_branchiost	477	27	100.0	154	2	Q8NAT4_HUMAN	Q8nat4_homo sapien
405	27	100.0	151	2	Q6XIO8_DROYA	Q6xio8_drosophila	478	27	100.0	154	2	Q869U7_DICDI	Q869u7_dictyosteli
406	27	100.0	151	2	Q8I3U6_PLAP7	Q8i3u6_plasmodium	479	27	100.0	154	2	Q8H639_ORYSA	Q8h639_oryza sativ
407	27	100.0	151	2	Q8WSQ7_PENJP	Q8wsq7_pemnaeus jap	480	27	100.0	154	2	Q72C75_DESVH	Q72c75_desulfovibr
408	27	100.0	151	2	Q962R5_SPOFR	Q962r5_stomodoptera	481	27	100.0	155	1	RS18_ENTHI	P48151 entamoeba h
409	27	100.0	151	2	Q9XYQ4_STOCA	Q9xyq4_stomoxys ca	482	27	100.0	155	2	Q5B1Y9_EMENI	Q5b1y9_aspergillus
410	27	100.0	151	2	Q4N2R9_THEPA	Q4n2r9_theileria p	483	27	100.0	155	2	Q5TU31_ANOGA	Q5tu31_anopheles g
411	27	100.0	151	2	Q4UA32_THEAN	Q4ua32_theileria a	484	27	100.0	155	2	Q56FE3_9HYME	Q56fe3_lysiphebus
412	27	100.0	151	2	Q4PM10_IXOSC	Q4pm10_ixodes scap	485	27	100.0	155	2	Q6ZFM3_ORYSA	Q6zfm3_oryza sativ
413	27	100.0	151	2	Q4GXR9_9CUCU	Q4gxr9_curculio gl	486	27	100.0	155	2	Q5SGZ1_THET8	Q5sgz1_thermus the
414	27	100.0	151	2	Q5Z552_ORYSA	Q5z552_oryza sativ	487	27	100.0	155	2	Q72HB9_THET2	Q72hb9_thermus the
415	27	100.0	151	2	Q6K6N6_ORYSA	Q6k6n6_oryza sativ	488	27	100.0	156	2	Q5EN04_MAGGR	Q5en04_magnaporthe
416	27	100.0	151	2	Q6ZA24_ORYSA	Q6za24_oryza sativ	489	27	100.0	156	2	Q527M9_MAGGR	Q527m9_magnaporthe
417	27	100.0	151	2	Q84OX5_ORYSA	Q84qx5_oryza sativ	490	27	100.0	156	2	Q6AW80_HUMAN	Q6aw80_homo sapien
418	27	100.0	151	2	Q7XSA6_ORYSA	Q7xsa6_oryza sativ	491	27	100.0	156	2	Q7R780_PLAYO	Q7r780_plasmodium
419	27	100.0	151	2	Q9F026_9PROT	Q9f026_unidentifie	492	27	100.0	156	2	Q8IIA2_PLAF7	Q8iaa2_plasmodium
420	27	100.0	151	2	Q70569_MOUSE	Q70569_mus musculu	493	27	100.0	156	2	Q4Z0S1_PLABE	Q4z0s1_plasmodium
421	27	100.0	151	2	Q544W4_MOUSE	Q544w4_mus musculu	494	27	100.0	156	2	Q4XNE9_PLACH	Q4xne9_plasmodium
422	27	100.0	151	2	Q6PDV6_RAT	Q6pdv6_rattus norv	495	27	100.0	156	2	Q9DL01_ORYSA	Q9dl01_oryza sativ
423	27	100.0	151	2	Q5DVH5_PLAFE	Q5dvh5_platichthys	496	27	100.0	157	2	Q5DE77_SCHJA	Q5de77_schistosoma
424	27	100.0	151	2	Q5ZHW8_CHICK	Q5zhw8_gallus gall	497	27	100.0	158	2	Q8YTP5_ORYSA	Q8ytp5_oryza sativ
425	27	100.0	151	2	Q6PBW3_BRABE	Q6pbw3_brachydanio	498	27	100.0	158	2	Q7XD88_ORYSA	Q7xd88_oryza sativ
426	27	100.0	151	2	Q6PI76_XENLA	Q6pi76_xenopus lae	499	27	100.0	158	2	Q8W375_ORYSA	Q8w375_oryza sativ
427	27	100.0	151	2	Q4SQ18_TETNG	Q4sq18_tetradodon n	500	27	100.0	159	2	Q74Z18_ASHGO	Q74z18_ashbya gos
428	27	100.0	152	1	RS14_CAEEL	P48150 caenorhabdi	501	27	100.0	159	2	Q76KS8_TRIVA	Q76ks8_trichomonas
429	27	100.0	152	1	RS18_AEQIR	Q8it98_aequiptecten	502	27	100.0	159	2	Q6EQK5_ORYSA	Q6eqk5_oryza sativ
430	27	100.0	152	1	RS18_BRABE	Q8lep0_branchiosteo	503	27	100.0	159	2	Q98S22_GUITH	Q98s22_guillardia
431	27	100.0	152	1	RS18_BRARE	Q8jgs9_brachydanio	504	27	100.0	160	2	Q6CAS5_YARLI	Q6cas5_yarrowia li
432	27	100.0	152	1	RS18_CANFA	Q5tfe9_canis famil	505	27	100.0	160	2	Q5VQE2_ORYSA	Q5vqe2_oryza sativ
433	27	100.0	152	1	RS18_DROME	P41094_drosophila	506	27	100.0	161	2	Q6ETF3_ORYSA	Q6etf3_oryza sativ
434	27	100.0	152	1	RS18_HUMAN	P62269_homo sapien	507	27	100.0	161	2	Q8LIS8_ORYSA	Q8lis8_oryza sativ
435	27	100.0	152	1	RS18_ICTPU	P62269_homo sapien	508	27	100.0	162	2	Q6ZCP7_ORYSA	Q6zcp7_oryza sativ
436	27	100.0	152	1	RS18_MOUSE	P62270_mus musculu	509	27	100.0	162	2	Q89900_9BETA	Q89900_human herpe
437	27	100.0	152	1	RS18_PIG	P62272_sus scrofa	510	27	100.0	163	2	Q6ZIH4_ORYSA	Q6zih4_oryza sativ
438	27	100.0	152	1	RS18_RAT	P62271_rattus norv	511	27	100.0	163	2	Q9RF18_9PROT	Q9rf18_amoeba anoxic
439	27	100.0	152	1	RS18_SPOFR	Q962r1_spodoptera	512	27	100.0	163	2	Q9CUJ5_MOUSE	Q9cujs_mus musculu
440	27	100.0	152	2	Q6ZV15_HUMAN	Q6zvi15_homo sapien	513	27	100.0	164	2	Q75LM6_ORYSA	Q75lm6_oryza sativ
441	27	100.0	152	2	Q5SUJ3_HUMAN	Q5suj3_homo sapien	514	27	100.0	164	2	Q8GRK5_ORYSA	Q8grk5_oryza sativ
442	27	100.0	152	2	Q5DAC3_SCHJA	Q5dac4_schistosoma	515	27	100.0	164	2	Q8GSL3_ORYSA	Q8gsl3_oryza sativ
443	27	100.0	152	2	Q5R269_DROSE	Q5r269_drosophila	516	27	100.0	165	2	Q6J7Z0_9VIRU	Q6j7z0_actinoplane
444	27	100.0	152	2	Q5R288_DROSI	Q5r288_drosophila	517	27	100.0	165	2	Q7XG68_ORYSA	Q7xg68_oryza sativ
445	27	100.0	152	2	Q5UAM4_BOMMO	Q5uam4_bombyx mori	518	27	100.0	165	2	Q8S6W9_ORYSA	Q8s6w9_oryza sativ
446	27	100.0	152	2	Q61MS9_CABER	Q61ms9_caenorhabdi	519	27	100.0	166	2	Q7QR09_GIALA	Q7qr09_giardia lam
447	27	100.0	152	2	Q6EUY1_9CUCU	Q6euy1_tinarcha ba	520	27	100.0	166	2	Q69SF9_ORYSA	Q69sf9_oryza sativ
448	27	100.0	152	2	Q6EUY1_9COLE	Q6euy2_dascillus c	521	27	100.0	166	2	Q6H410_ORYSA	Q6h410_oryza sativ
449	27	100.0	152	2	Q6EUY3_CICCA	Q6euy3_plutella xy	522	27	100.0	166	2	Q7XG60_ORYSA	Q7xg60_oryza sativ
450	27	100.0	152	2	Q6F458_PLUXY	Q6f458_plutella c	523	27	100.0	166	2	Q8S5G2_ORYSA	Q8s5g2_oryza sativ
451	27	100.0	152	2	Q616G6_ANTYA	Q616g6_antheraea y	524	27	100.0	167	2	Q564T3_CAEEL	Q564t3_caenorhabdi
452	27	100.0	152	2	Q6XHV8_DROYA	Q6xhv8_drosophila	525	27	100.0	167	2	Q5Z4K7_ORYSA	Q5z4k7_oryza sativ
453	27	100.0	152	2	Q7PRH3_ANOGA	Q7prh3_anopheles g	526	27	100.0	167	2	Q7NDM1_GLOVI	Q7ndm1_gloeobacter
454	27	100.0	152	2	Q7QBX2_ANOGA	Q7qbx2_anopheles g	527	27	100.0	169	2	Q6H7Y8_ORYSA	Q6h7y8_oryza sativ
455	27	100.0	152	2	Q7QEH1_ANOGA	Q7qeh1_anopheles g	528	27	100.0	169	2	Q6YVF5_ORYSA	Q6yvf5_oryza sativ
456	27	100.0	152	2	Q7YU94_IXORI	Q7yuy94_ixodes ric	529	27	100.0	169	2	Q57E77_BRUAB	Q57e77_brucella ab
457	27	100.0	152	2	Q8MT19_DROME	Q8mt19_drosophila	530	27	100.0	169	2	Q8G1P7_BRUSU	Q8g1p7_brucella su
458	27	100.0	152	2	Q4KTC3_SURDO	Q4ktc3_suberites d	531	27	100.0	169	2	Q8YG77_BRUME	Q8yg77_brucella me
459	27	100.0	152	2	Q4H450_CRAGI	Q4h450_crassostrea	532	27	100.0	170	2	Q6YVK9_ORYSA	Q6yvk9_oryza sativ
460	27	100.0	152	2	Q4GKQ1_9CUCU	Q4gxq1_curculio gl	533	27	100.0	170	2	Q6YKJ4_ORYSA	Q6ywj4_oryza sativ
461	27	100.0	152	2	Q4GXP9_9COLE	Q4gxp9_sphaerius s	534	27	100.0	170	2	Q6AUV4_ORYSA	Q6auv4_oryza sativ
462	27	100.0	152	2	Q5VQG5_ORYSA	Q5vqg5_oryza sativ	535	27	100.0	170	2	Q8YXT9_ANASP	Q8yxt9_anabaena sp
463	27	100.0	152	2	Q5ZE43_ORYSA	Q5ze43_oryza sativ	536	27	100.0	170	2	Q72A92_DESVH	Q72a92_desulfovibr
464	27	100.0	152	2	Q8VK26_MYCTU	Q8vk26_mycobacteri	537	27	100.0	171	2	Q8GRX0_ORYSA	Q8grx0_oryza sativ
465	27	100.0	152	2	Q7U062_MYCBO	Q7u062_mycobacteri	538	27	100.0	172	2	Q8IP68_DROME	Q8ip68_drosophila
466	27	100.0	152	2	Q9R1Z3_STRCO	Q9riz3_streptomyce	539	27	100.0	172	2	Q6EPX4_ORYSA	Q6epx4_oryza sativ
467	27	100.0	152	2	P71802_MYCTU	P71802_mycobacteri	540	27	100.0	172	2	Q4NZE3_9BELT	Q4nze3_aeromonas xob
468	27	100.0	152	2	Q561N5_MOUSE	Q561n5_mus musculu	541	27	100.0	172	2	Q9A3H2_CAUCR	Q9a3h2_caulobacter
469	27	100.0	152	2	Q5SVZ5_MOUSE	Q5svz5_mus musculu	542	27	100.0	174	2	Q52DI3_MAGGR	Q52di3_magnaporthe

543	27	100.0	174	2	Q4J5A8_AZOV1	Q4J5A8	azotobacter	616	27	100.0	208	2	Q5V5K6_HALMA	Q5V5K6	haloarcula
544	27	100.0	174	2	Q8FTX3_COREF	Q8FTX3	corynebacte	617	27	100.0	208	2	Q5ZD03_ORYSA	Q5ZD03	oryza sativ
545	27	100.0	175	2	Q6YUN1_ORYSA	Q6YUN1	oryza sativ	618	27	100.0	209	2	Q5WMT0_ORYSA	Q5WMT0	oryza sativ
546	27	100.0	175	2	Q6FFN0_ACTAD	Q6FFN0	acinetobact	619	27	100.0	209	2	Q8H5G7_ORYSA	Q8H5G7	oryza sativ
547	27	100.0	176	2	Q84M52_ORYSA	Q84M52	oryza sativ	620	27	100.0	209	2	Q8H5G7_ORYSA	Q8H5G7	oryza sativ
548	27	100.0	176	2	Q5S3R2_9BACT	Q5S3R2	uncultured	621	27	100.0	210	2	Q6ZVB9_HUMAN	Q6ZVB9	homo sapien
549	27	100.0	177	2	Q6ORS1_CAEBR	Q6ORS1	caenorhabdi	622	27	100.0	210	2	Q8LHV7_ORYSA	Q8LHV7	oryza sativ
550	27	100.0	177	2	Q5N7U3_ORYSA	Q5N7U3	oryza sativ	623	27	100.0	210	2	Q8LHV7_ORYSA	Q8LHV7	oryza sativ
551	27	100.0	177	2	Q5JL76_ORYSA	Q5JL76	oryza sativ	624	27	100.0	210	2	Q8LHV7_ORYSA	Q8LHV7	oryza sativ
552	27	100.0	177	2	Q5QN35_ORYSA	Q5QN35	oryza sativ	625	27	100.0	211	2	Q4UD17_THEAN	Q4UD17	theileria a
553	27	100.0	177	2	Q5SMQ7_ORYSA	Q5SMQ7	oryza sativ	626	27	100.0	211	2	Q6ZB01_ORYSA	Q6ZB01	oryza sativ
554	27	100.0	177	2	Q88H82_PSEPK	Q88H82	pseudomonas	627	27	100.0	211	2	Q8H7Z9_ORYSA	Q8H7Z9	oryza sativ
555	27	100.0	177	2	Q7NJ48_GLOVI	Q7NJ48	gloeobacter	628	27	100.0	213	2	Q6C992_YARLI	Q6C992	yarrowia li
556	27	100.0	178	2	Q7XTX4_ORYSA	Q7XTX4	oryza sativ	629	27	100.0	213	2	Q6K5M0_ORYSA	Q6K5M0	oryza sativ
557	27	100.0	179	2	Q417B5_GIBZE	Q417B5	gibberella	630	27	100.0	213	2	Q6AES1_LEIXX	Q6AES1	leifsonia x
558	27	100.0	179	2	Q5DF46_SCHJA	Q5DF46	schistosoma	631	27	100.0	213	2	Q8FSA4_COREF	Q8FSA4	corynebacte
559	27	100.0	180	2	Q8N7E7_HUMAN	Q8N7E7	homo sapien	632	27	100.0	213	2	Q800F5_TETNG	Q800F5	tetraodon n
560	27	100.0	181	2	Q6ORQ2_ORYSA	Q6ORQ2	oryza sativ	633	27	100.0	214	2	Q4TE23_TETNG	Q4TE23	tetraodon n
561	27	100.0	182	2	Q5VUB1_HUMAN	Q5VUB1	homo sapien	634	27	100.0	216	2	Q8XOV9_NEUCR	Q8XOV9	neurospora
562	27	100.0	182	2	Q53HY0_LUPAL	Q53HY0	lupinus alb	635	27	100.0	216	2	Q8H351_ORYSA	Q8H351	oryza sativ
563	27	100.0	182	2	Q9A9B0_CAUCR	Q9A9B0	caulobacter	636	27	100.0	216	2	Q9M293_ARATH	Q9M293	arabidopsis
564	27	100.0	183	2	Q653R7_ORYSA	Q653R7	oryza sativ	637	27	100.0	216	2	Q8ZHB8_YERPE	Q8ZHB8	versinia pe
565	27	100.0	183	2	Q6H5L5_ORYSA	Q6H5L5	oryza sativ	638	27	100.0	216	2	Q666J9_YERPS	Q666J9	versinia ps
566	27	100.0	183	2	Q9A9T3_CAUCR	Q9A9T3	caulobacter	639	27	100.0	216	2	Q74DM3_GEOSL	Q74DM3	geobacter s
567	27	100.0	184	2	Q9ALU5_ORYSA	Q9ALU5	oryza sativ	640	27	100.0	217	2	Q69N81_ORYSA	Q69N81	oryza sativ
568	27	100.0	184	2	Q8H5M9_ORYSA	Q8H5M9	oryza sativ	641	27	100.0	217	2	Q8S1Q1_ORYSA	Q8S1Q1	oryza sativ
569	27	100.0	184	2	Q7NNM5_GLOVI	Q7NNM5	gloeobacter	642	27	100.0	217	2	Q5F6Y1_NEIG1	Q5F6Y1	neisseria g
570	27	100.0	185	2	Q5Z849_ORYSA	Q5Z849	oryza sativ	643	27	100.0	217	2	Q5F6Z5_NEIG1	Q5F6Z5	neisseria g
571	27	100.0	185	2	Q8KH53_PARPV	Q8KH53	paracoccu	644	27	100.0	217	2	Q5F769_NEIG1	Q5F769	neisseria g
572	27	100.0	185	2	Q5G267_XANOR	Q5G267	xanthomonas	645	27	100.0	217	2	Q5F781_NEIG1	Q5F781	neisseria g
573	27	100.0	185	2	Q80OG0_TETNG	Q80OG0	tetraodon n	646	27	100.0	217	2	Q5F7C2_NEIG1	Q5F7C2	neisseria g
574	27	100.0	186	2	Q859T5_9VIRU	Q859T5	bacterioph	647	27	100.0	217	2	Q5F7G7_NEIG1	Q5F7G7	neisseria g
575	27	100.0	186	2	Q5QN63_ORYSA	Q5QN63	oryza sativ	648	27	100.0	217	2	Q5F7L4_NEIG1	Q5F7L4	neisseria g
576	27	100.0	188	2	Q8NK58_PARRR	Q8NK58	paracoccidi	649	27	100.0	217	2	Q5F971_NEIG1	Q5F971	neisseria g
577	27	100.0	189	1	RAS1_DROMA	RAS1	drosophila	650	27	100.0	217	2	Q5FA33_NEIG1	Q5FA33	neisseria g
578	27	100.0	189	1	RAS1_DROME	RAS1	drosophila	651	27	100.0	217	2	Q9AH65_NEIGO	Q9AH65	neisseria g
579	27	100.0	189	1	RAS1_DROSI	RAS1	drosophila	652	27	100.0	217	2	Q7BLD1_NEIME	Q7BLD1	neisseria m
580	27	100.0	189	2	Q81GM6_DROME	Q81GM6	drosophila	653	27	100.0	217	2	Q5F549_NEIG1	Q5F549	neisseria g
581	27	100.0	189	2	Q652S9_ORYSA	Q652S9	oryza sativ	654	27	100.0	217	2	Q5F8T7_NEIG1	Q5F8T7	neisseria g
582	27	100.0	189	2	Q7X9L7_WHEAT	Q7X9L7	triticultum ae	655	27	100.0	217	2	Q9JUSQ4_NEIMA	Q9JUSQ4	neisseria m
583	27	100.0	190	2	Q9A387_CAUCR	Q9A387	caulobacter	656	27	100.0	217	2	Q9JVN3_NEIMA	Q9JVN3	neisseria m
584	27	100.0	192	2	Q6Z947_ORYSA	Q6Z947	oryza sativ	657	27	100.0	217	2	Q9JWA4_NEIMA	Q9JWA4	neisseria m
585	27	100.0	192	2	Q8KVX2_9RHOB	Q8KVX2	ruegeria sp	658	27	100.0	217	2	Q9JYV8_NEIMB	Q9JYV8	neisseria m
586	27	100.0	193	2	Q5Z7E3_ORYSA	Q5Z7E3	oryza sativ	659	27	100.0	218	2	Q6ZV46_HUMAN	Q6ZV46	homo sapien
587	27	100.0	195	2	Q5VMU5_ORYSA	Q5VMU5	oryza sativ	660	27	100.0	218	2	Q6YV93_ORYSA	Q6YV93	oryza sativ
588	27	100.0	195	2	Q5VQ95_ORYSA	Q5VQ95	oryza sativ	661	27	100.0	219	2	Q8TW07_METKA	Q8TW07	methanopyru
589	27	100.0	195	2	Q6EP78_ORYSA	Q6EP78	oryza sativ	662	27	100.0	219	2	Q5ZHA0_9VIRU	Q5ZHA0	torque teno
590	27	100.0	196	2	Q6ZG43_ORYSA	Q6ZG43	oryza sativ	663	27	100.0	220	2	Q9C8A1_ARATH	Q9C8A1	arabidopsis
591	27	100.0	196	2	Q4SMD6_TETNG	Q4SMD6	tetraodon n	664	27	100.0	221	2	Q9AD79_STRCO	Q9AD79	streptomyce
592	27	100.0	197	2	Q6ESG5_ORYSA	Q6ESG5	oryza sativ	665	27	100.0	222	2	Q7QQ11_GIALA	Q7QQ11	giardia lam
593	27	100.0	197	2	Q8UE74_AGR5	Q8UE74	agrobacteri	666	27	100.0	222	2	Q5VP19_ORYSA	Q5VP19	oryza sativ
594	27	100.0	197	2	Q9JU54_NEIMA	Q9JU54	neisseria m	667	27	100.0	222	2	Q6EUQ3_ORYSA	Q6EUQ3	oryza sativ
595	27	100.0	197	2	Q9JW43_NEIMA	Q9JW43	neisseria m	668	27	100.0	222	2	Q9JT64_NEIMA	Q9JT64	neisseria m
596	27	100.0	198	2	Q9Y8S5_AERPE	Q9Y8S5	aeropyrum p	669	27	100.0	222	2	Q9KOLL_NEIMB	Q9KOLL	neisseria m
597	27	100.0	198	2	Q4HXAL_GIBZE	Q4HXAL	gibberella	670	27	100.0	223	2	Q7XVX9_ORYSA	Q7XVX9	oryza sativ
598	27	100.0	198	2	Q84T32_ORYSA	Q84T32	oryza sativ	671	27	100.0	223	2	Q80Z11_MOUSE	Q80Z11	mus musculus
599	27	100.0	198	2	Q5YS59_NOCFA	Q5YS59	nocardia fa	672	27	100.0	224	1	HXB6B_BRARE	HXB6B	brachydanio
600	27	100.0	198	2	Q4SJK7_TETNG	Q4SJK7	tetraodon n	673	27	100.0	224	1	HXB6_HUMAN	HXB6	homo sapien
601	27	100.0	199	2	Q5JND0_ORYSA	Q5JND0	oryza sativ	674	27	100.0	224	1	HXB6_MOUSE	HXB6	mouse
602	27	100.0	199	2	Q6Z6U7_ORYSA	Q6Z6U7	oryza sativ	675	27	100.0	224	2	Q7CVA9_AGR5	Q7CVA9	agrobacteri
603	27	100.0	200	2	Q9JU18_NEIMA	Q9JU18	neisseria m	676	27	100.0	225	2	Q6Z0Y6_ORYSA	Q6Z0Y6	oryza sativ
604	27	100.0	200	2	Q8UY85_9ADEN	Q8UY85	simian aden	677	27	100.0	225	2	Q53IA3_STRGR	Q53IA3	streptomyce
605	27	100.0	201	2	Q9LHX7_ORYSA	Q9LHX7	oryza sativ	678	27	100.0	225	2	Q4NX55_9DELT	Q4NX55	anaeromyxob
606	27	100.0	202	2	Q61ZS3_CAEBR	Q61ZS3	caenorhabdi	679	27	100.0	226	2	Q5FSP2_GLUOX	Q5FSP2	gluconobact
607	27	100.0	202	2	Q859U0_9VIRU	Q859U0	bacterioph	680	27	100.0	227	1	RS3_METKA	RS3	methanopyru
608	27	100.0	202	2	Q8DD32_VIBVU	Q8DD32	vibrio vuln	681	27	100.0	227	1	US08_HCMVA	US08	human cytom
609	27	100.0	205	2	Q7RXK0_NEUCR	Q7RXK0	neurospora	682	27	100.0	227	2	Q6NTR9_HUMAN	Q6NTR9	homo sapien
610	27	100.0	205	2	Q6H404_ORYSA	Q6H404	oryza sativ	683	27	100.0	227	2	Q7FAC3_ORYSA	Q7FAC3	oryza sativ
611	27	100.0	205	2	Q8GVG8_ORYSA	Q8GVG8	oryza sativ	684	27	100.0	227	2	Q7X6T6_ORYSA	Q7X6T6	oryza sativ
612	27	100.0	206	2	Q6ZAF4_ORYSA	Q6ZAF4	oryza sativ	685	27	100.0	227	2	Q6SVZ8_HCMV	Q6SVZ8	human cytom
613	27	100.0	207	1	3MGH_XANCP	Q8P8C7	xanthomonas	686	27	100.0	228	1	HXB6A_BRARE	HXB6A	brachydanio
614	27	100.0	207	2	Q5W6D3_ORYSA	Q5W6D3	oryza sativ	687	27	100.0	228	2	Q8K184_MOUSE	Q8K184	mus musculus
615	27	100.0	207	2	Q4UVR0_XANCP	Q4UVR0	xanthomonas	688	27	100.0	228	2	Q9CZ37_MOUSE	Q9CZ37	mus musculus

689	27	100.0	229	1	HXA6_HETFR	Q9ia24 heterodontu	762	27	100.0	256	2	Q8UFW3_AGRt5	Q8Ufw3 agrobacteri
690	27	100.0	229	2	Q9BWT5_HUMAN	Q9bwt5 homo sapien	763	27	100.0	257	1	CJ095_HUMAN	Q9N7t3 homo sapien
691	27	100.0	229	2	Q4H5C8_9DEIO	Q4h5c8 deinococcus	764	27	100.0	258	2	Q9R9LI_RHIME	Q9r9l1 rhizobium m
692	27	100.0	230	2	Q6Z5Q5_ORYSA	Q6z5q5 oryza sativ	765	27	100.0	257	2	Q7X1C7_9BACT	Q7x1c7 leptospiril
693	27	100.0	231	1	HXA6_CHICK	Q5ylh5 gallus gall	766	27	100.0	258	2	Q5YS46_NOCFA	Q5ys46 nocardia fa
694	27	100.0	231	2	Q69LJ5_ORYSA	Q69lj5 oryza sativ	767	27	100.0	258	2	Q82E81_STRAW	Q82e81 streptomyce
695	27	100.0	231	2	Q9ZUR1_ARATH	Q9zur1 arabidopsis	768	27	100.0	259	2	Q4I8R8_GIBZE	Q4i8r8 gibberella
696	27	100.0	231	2	Q8LHQ0_ORYSA	Q8lhq0 oryza sativ	769	27	100.0	259	2	Q4IJX7_GIBZE	Q4ijx7 gibberella
697	27	100.0	232	1	Q63UD5_BURPS	Q63ud5 burkholderi	770	27	100.0	259	2	Q4WAI8_ASPFU	Q4wai8 aspergillus
698	27	100.0	232	1	HXA6_MOUSE	P09092 mus musculu	771	27	100.0	259	2	Q5VMF8_ORYSA	Q5vmf8 oryza sativ
699	27	100.0	232	1	Q5P8U8_AZOSE	Q5p8u8 azoarcus sp	772	27	100.0	260	2	Q96IL7_HUMAN	Q96il7 homo sapien
700	27	100.0	233	1	HXA6_HUMAN	P31267 homo sapien	773	27	100.0	260	2	Q9FMU3_ORYSA	Q9fmw3 oryza sativ
701	27	100.0	233	2	Q00525_CLAPU	Q00525 claviceps p	774	27	100.0	260	2	Q9FOY2_STRCO	Q9fow2 streptomyce
702	27	100.0	233	2	Q6YI67_ORYSA	Q6y677 oryza sativ	775	27	100.0	260	2	Q8ZQN5_STRAW	Q8zqn5 streptomyce
703	27	100.0	234	2	Q6W985_9PERC	Q6w985 spheroeoides	776	27	100.0	261	2	Q9BSG5_HUMAN	Q9bsg5 homo sapien
704	27	100.0	234	2	Q5VU80_HUMAN	Q5vu80 homo sapien	777	27	100.0	261	2	Q5KQH9_ORYSA	Q5kqh9 oryza sativ
705	27	100.0	234	2	Q8I747_SUBDO	Q8i747 suberites d	778	27	100.0	261	2	Q69WY0_ORYSA	Q69wy0 oryza sativ
706	27	100.0	234	2	Q625R6_ORYSA	Q625r6 oryza sativ	779	27	100.0	261	2	Q9FYA8_ARATH	Q9fyar arabidopsis
707	27	100.0	235	2	Q5VQJ3_ORYSA	Q5vqj3 oryza sativ	780	27	100.0	262	2	Q652G3_ORYSA	Q652g3 oryza sativ
708	27	100.0	235	2	Q76ZAB_ORYSA	Q76zab oryza sativ	781	27	100.0	262	2	Q652Z0_ORYSA	Q652z0 oryza sativ
709	27	100.0	235	2	Q4NML0_9MITC	Q4nml0 arthrobacte	782	27	100.0	262	2	Q7XUT4_ORYSA	Q7xut4 oryza sativ
710	27	100.0	236	1	UL51_PRVKA	Q85227 pseudorabie	783	27	100.0	262	2	Q9LHP2_ARATH	Q9lhp2 arabidopsis
711	27	100.0	236	1	Q8L7Q4_ARATH	Q8l7q4 arabidopsis	784	27	100.0	263	2	Q6ZLW9_ORYSA	Q6zlw9 oryza sativ
712	27	100.0	236	2	Q5YX13_NOCFA	Q5yx13 nocardia fa	785	27	100.0	263	2	Q949P4_ARATH	Q949p4 arabidopsis
713	27	100.0	236	2	Q5PP96_9ALPH	Q5pp96 suid herpes	786	27	100.0	263	2	Q7UGT3_RHOBA	Q7ugt3 rhodopirell
714	27	100.0	237	2	Q53MG2_ORYSA	Q53mg2 oryza sativ	787	27	100.0	264	2	Q96C55_HUMAN	Q96c55 homo sapien
715	27	100.0	238	2	Q8TYE7_METKA	Q8tye7 methanopyru	788	27	100.0	264	2	Q6NW31_HUMAN	Q6nw31 homo sapien
716	27	100.0	238	2	Q6ZAJ5_ORYSA	Q6zaj5 oryza sativ	789	27	100.0	264	2	Q7XGE8_ORYSA	Q7xge8 oryza sativ
717	27	100.0	238	2	Q94DG7_ORYSA	Q94dg7 oryza sativ	790	27	100.0	264	2	Q940D9_PINTA	Q940d9 pinus taeda
718	27	100.0	238	2	Q4TF80_TETNG	Q4tf80 tetraodon n	791	27	100.0	265	2	Q5VU79_HUMAN	Q5vu79 homo sapien
719	27	100.0	239	2	Q4NS86_9DELT	Q4ns86 anaeromyxob	792	27	100.0	265	2	Q77230_BRUPA	Q77230 brugia paha
720	27	100.0	239	2	Q9S287_STRCO	Q9s287 streptomyce	793	27	100.0	265	2	Q4PLW3_DROME	Q4plw3 drosophila
721	27	100.0	239	2	Q9ANJ3_BRAJA	Q9anj3 bradythizob	794	27	100.0	265	2	Q5JLCA_ORYSA	Q5jlc4 oryza sativ
722	27	100.0	240	1	BIRA_PARDE	P29906 paracoccus	795	27	100.0	265	2	Q69PF4_ORYSA	Q69pf4 oryza sativ
723	27	100.0	241	2	Q69N84_ORYSA	Q69n84 oryza sativ	796	27	100.0	265	2	Q75KB6_ORYSA	Q75kb6 oryza sativ
724	27	100.0	241	2	Q4NPS6_9DELT	Q4np6 anaeromyxob	797	27	100.0	265	2	Q57DT7_BRUAB	Q57dt7 brucella ab
725	27	100.0	242	2	Q53PT8_ORYSA	Q53pt8 oryza sativ	798	27	100.0	265	2	Q8G1A3_BRUSU	Q8g1a3 brucella su
726	27	100.0	242	2	Q52660_RHOCA	Q52660 rhodobacter	799	27	100.0	265	2	Q914K0_9VIRU	Q914k0 sulfolobus
727	27	100.0	242	2	Q82RL6_STRAW	Q82rl6 streptomyce	800	27	100.0	266	2	Q6K3Q7_ORYSA	Q6k3q7 oryza sativ
728	27	100.0	242	2	Q4RUF0_TETNG	Q4ruf0 tetraodon n	801	27	100.0	266	2	Q62B47_BURMA	Q62b47 burkholderi
729	27	100.0	243	2	Q52G27_MAGGR	Q52g27 magnaporthe	802	27	100.0	266	2	Q5ENX6_TORQUE	Q5enx6 torque teno
730	27	100.0	243	2	Q7F6X1_ORYSA	Q7f6x1 oryza sativ	803	27	100.0	267	2	Q8PSE4_METMA	Q8pse4 methanosarc
731	27	100.0	243	2	Q6AHG3_LEIXX	Q6ahg3 leifsonia x	804	27	100.0	267	2	Q4X1D3_ASPFU	Q4xld3 aspergillus
732	27	100.0	244	2	Q61146_DROME	Q61146 drosophila	805	27	100.0	267	2	Q84SX4_ORYSA	Q84sx4 oryza sativ
733	27	100.0	245	1	C1OA_HUMAN	P02745 homo sapien	806	27	100.0	267	2	Q56JH5_BRARE	Q56jh5 brachydanio
734	27	100.0	245	2	Q62M03_BURMA	Q62m03 burkholderi	807	27	100.0	268	2	Q4NN68_9DELT	Q4nn68 anaeromyxob
735	27	100.0	246	2	Q9Z9I3_STRFR	Q9z9i3 streptomyce	808	27	100.0	269	2	Q8T0C2_DROME	Q8t0c2 drosophila
736	27	100.0	246	2	Q4NS99_9DELT	Q4ne99 anaeromyxob	809	27	100.0	269	2	Q5P2L0_AZOSE	Q5p2l0 azoarcus sp
737	27	100.0	246	2	P72243_PPRERU	P72243 prevotella	810	27	100.0	270	2	Q8H597_ORYSA	Q8h597 oryza sativ
738	27	100.0	246	2	Q6NA33_RHOFA	Q6na33 rhodopseudo	811	27	100.0	271	2	Q8XOM0_9STRA	Q8xom0 phytophthor
739	27	100.0	246	2	Q5ZMA3_CHICK	Q5zma3 gallus gall	812	27	100.0	272	2	Q753B0_ASHGO	Q753b0 ashbya goss
740	27	100.0	247	2	Q8N1W5_HUMAN	Q8n1w5 homo sapien	813	27	100.0	272	2	Q6ZG36_ORYSA	Q6z36 oryza sativ
741	27	100.0	247	2	Q8LMM3_ORYSA	Q8lmm3 oryza sativ	814	27	100.0	272	2	Q7XGE9_ORYSA	Q7xge9 oryza sativ
742	27	100.0	247	2	Q5ZAH1_ORYSA	Q5zah1 oryza sativ	815	27	100.0	272	2	Q8H7C6_ARATH	Q8h7c6 arabidopsis
743	27	100.0	247	2	Q4UWH3_XANCP	Q4uwh3 xanthomonas	816	27	100.0	272	2	Q86624_STRCO	Q86624 streptomyce
744	27	100.0	247	2	Q8E7M7_XANCP	Q8e7m7 xanthomonas	817	27	100.0	272	2	Q62N03_BURMA	Q62n03 burkholderi
745	27	100.0	248	2	Q7QYI1_GITALA	Q7qyi1 giardia lam	818	27	100.0	273	2	Q9NUN8_HUMAN	Q9nun8 homo sapien
746	27	100.0	248	2	Q67UX6_ORYSA	Q67ux6 oryza sativ	819	27	100.0	273	2	Q7X4A0_9CYAN	Q7x4a0 hapalosipho
747	27	100.0	249	2	Q611L1_DROME	Q611l1 drosophila	820	27	100.0	273	2	Q7X4A0_9CYAN	Q7x4a0 tolypotherix
748	27	100.0	249	2	Q5LPT7_SILPO	Q5lpt7 silicibacte	821	27	100.0	273	2	Q8L130_9CYAN	Q8l130 chroococcid
749	27	100.0	250	2	Q9A6Z0_CAUCR	Q9a6z0 caulobacter	822	27	100.0	273	2	Q8L132_9CYAN	Q8l132 pleurocappa
750	27	100.0	250	2	Q73U47_MYCPA	Q73u47 mycobacteri	823	27	100.0	273	2	Q8L136_9CYAN	Q8l136 xenococcus
751	27	100.0	251	2	Q5GX51_XANOR	Q5gx51 xanthomonas	824	27	100.0	273	2	Q8L137_9CYAN	Q8l137 oscillatori
752	27	100.0	252	2	Q3O850_RABIT	Q3o850 cryptotlagus	825	27	100.0	273	2	Q98KS0_RHILO	Q98ks0 rhizobium l
753	27	100.0	252	2	Q521H7_NOCFA	Q521h7 nocardia fa	826	27	100.0	273	2	Q98194_9SYLV	Q98194 acrocephalu
754	27	100.0	253	2	Q84QML_ORYSA	Q84qm1 oryza sativ	827	27	100.0	274	2	Q5Z236_NOCFA	Q5z236 nocardia fa
755	27	100.0	253	2	Q6XN02_RHOER	Q6xn02 rhodococcus	828	27	100.0	276	2	Q8WV14_HUMAN	Q8wv14 homo sapien
756	27	100.0	254	2	Q5ZAL7_MAGGR	Q5zal7 magnaporthe	829	27	100.0	276	2	Q9CX12_MOUSE	Q9cx12 mus musculu
757	27	100.0	254	2	Q7XFD3_ORYSA	Q7xrd3 oryza sativ	830	27	100.0	277	2	Q7QCJ6_ANOGA	Q7qcj6 anopheles g
758	27	100.0	254	2	Q9AUN1_ORYSA	Q9aun1 oryza sativ	831	27	100.0	277	2	Q94LG8_ORYSA	Q94lg8 oryza sativ
759	27	100.0	255	2	Q75H82_ORYSA	Q75h82 oryza sativ	832	27	100.0	277	2	Q7XJ93_ARATH	Q7xj93 arabidopsis
760	27	100.0	255	2	Q9ZON8_RHIME	Q9zom8 rhizobium m	833	27	100.0	277	2	Q938E4_MYCSM	Q938e4 mycobacteri
761	27	100.0	255	2	Q7NNC7_GLOVI	Q7nnc7 gleobacter	834	27	100.0	277	2	Q90Y82_LAMJA	Q90y82 lampetra ja

835	27	100.0	278	2	Q62VD1_HUMAN	Q62vd1 homo sapien	908	27	100.0	295	2	Q8L148_9CYAN	Q8L148 calothrix b
836	27	100.0	278	2	Q57OR7_ARATH	Q57or7 arabidopsis	909	27	100.0	295	2	Q8L149_OSCAG	Q8L149 oscillatori
837	27	100.0	278	2	Q5FLK6_ARATH	Q5flk6 arabidopsis	910	27	100.0	295	2	Q8L150_ANAVA	Q8L150 anabaena va
838	27	100.0	278	2	Q4LYG8_9BURK	Q4lyg8 burkholderi	911	27	100.0	295	2	Q8L151_SPIPL	Q8L151 spirulina p
839	27	100.0	278	2	Q7VPJ1_HARDU	Q7vpj1 haemophilus	912	27	100.0	295	2	Q63PH1_BURPS	Q63ph1 burkholderi
840	27	100.0	278	2	Q9A295_CAUCR	Q9a295 caulobacter	913	27	100.0	295	2	Q62FS5_BURMA	Q62fs5 burkholderi
841	27	100.0	278	2	Q913X7_PSGAB	Q913x7 pseudomonas	914	27	100.0	296	2	Q6AV84_ORYSA	Q6av84 oryza sativ
842	27	100.0	279	2	Q6CD78_YARLI	Q6cd78 yarowia li	915	27	100.0	296	2	Q8L128_9SYNE	Q8L128 synecococc
843	27	100.0	279	2	Q4VWX2_HUMAN	Q4vwx2 homo sapien	916	27	100.0	297	2	Q4LZ46_9BURK	Q4lz46 burkholderi
844	27	100.0	279	2	Q4H9X4_9DEIO	Q4h9x4 deinococcus	917	27	100.0	298	2	Q74Z87_MYCPA	Q74z87 mycobacteri
845	27	100.0	280	2	Q5VU78_HUMAN	Q5vu78 homo sapien	918	27	100.0	300	2	Q92PC3_RHIME	Q92pc3 rhizobium m
846	27	100.0	280	2	Q4VOZ2_BACZZ	Q4voz2 bacillus ce	919	27	100.0	301	2	Q4S6R9_TETNG	Q4s6r9 tetraodon n
847	27	100.0	280	2	Q63XY7_BURPS	Q63xy7 burkholderi	920	27	100.0	302	1	BIRA_HAEIN	P46363 haemophilus
848	27	100.0	281	2	Q62BE3_CAENR	Q62be3 caenorhabdi	921	27	100.0	302	2	Q5JUG1_ORYSA	Q5j191 oryza sativ
849	27	100.0	281	2	Q7XMX3_ORYSA	Q7xmx3 oryza sativ	922	27	100.0	302	2	Q4QW6_HAB18	Q4qwm6 haemophilus
850	27	100.0	281	2	Q4TT17_9SPHN	Q4tt17 erythrobact	923	27	100.0	303	2	Q59GR8_HUMAN	Q59gr8 homo sapien
851	27	100.0	282	2	Q8T8R8_DROME	Q8t8r8 drosophila	924	27	100.0	303	2	Q96I32_HUMAN	Q96i32 homo sapien
852	27	100.0	282	2	Q9VFL2_DROME	Q9vfl2 drosophila	925	27	100.0	303	2	Q8H5K7_ORYSA	Q8h5k7 oryza sativ
853	27	100.0	282	2	Q6S6E1_ORYSA	Q6s6e1 oryza sativ	926	27	100.0	304	2	Q4I643_GIBZE	Q4i643 gibberella
854	27	100.0	282	2	Q73U61_MYCPA	Q73u61 mycobacteri	927	27	100.0	304	2	Q4V3S5_DROME	Q4v3s5 drosophila
855	27	100.0	282	2	Q91TR4_TUHV1	Q91tr4 tupaiid her	928	27	100.0	305	2	Q6K9D9_ORYSA	Q6k9d9 oryza sativ
856	27	100.0	283	2	Q4WEZ3_ASPFU	Q4wez3 aspergillus	929	27	100.0	305	2	Q9H7Z4_HUMAN	Q9h7z4 homo sapien
857	27	100.0	283	2	Q6UUA4_ORYSA	Q6uu44 oryza sativ	930	27	100.0	305	2	Q4THN2_TETNG	Q4thn2 tetraodon n
858	27	100.0	283	2	Q84WF3_ORYSA	Q84wf3 oryza sativ	931	27	100.0	306	2	Q84S67_ORYSA	Q84s67 oryza sativ
859	27	100.0	283	2	Q8AVR7_XENLA	Q8avr7 xenopus lae	932	27	100.0	306	2	Q4SET5_TETNG	Q4set5 tetraodon n
860	27	100.0	285	2	Q4HUM8_GIBZE	Q4hum8 gibberella	933	27	100.0	307	2	Q6I3G0_CABBR	Q6i3g0 caenorhabdi
861	27	100.0	285	2	Q8IXW9_HUMAN	Q8ixw9 homo sapien	934	27	100.0	307	2	Q18327_CAEEL	Q18327 caenorhabdi
862	27	100.0	285	2	Q4Q2S4_LEIMA	Q4q2s4 leishmania	935	27	100.0	307	2	Q5NTH3_ORYSA	Q5nth3 oryza sativ
863	27	100.0	285	2	Q9U2U0_CABEL	Q9u2u0 caenorhabdi	936	27	100.0	307	2	Q5YS53_NOCFA	Q5ys53 nocardia fa
864	27	100.0	285	2	Q6YU24_ORYSA	Q6yu24 oryza sativ	937	27	100.0	309	2	Q5SKU7_CRYNE	Q5sku7 cryptococcu
865	27	100.0	286	2	Q9VGD1_DROME	Q9vgd1 drosophila	938	27	100.0	309	2	Q5KANI_CRYNE	Q5kan1 cryptococcu
866	27	100.0	287	2	Q53JU3_ORYSA	Q53ju3 oryza sativ	939	27	100.0	310	2	Q96KNI_HUMAN	Q96kn1 homo sapien
867	27	100.0	287	2	Q9SEU4_ARATH	Q9seu4 arabidopsis	940	27	100.0	310	2	Q62LF3_BURMA	Q62lf3 burkholderi
868	27	100.0	287	2	Q8YGL4_BRUME	Q8ygl4 brucella me	941	27	100.0	311	2	Q7NSC2_CHRVO	Q7nsc2 chroomobacte
869	27	100.0	288	1	IMPA2_HUMAN	Q14732 homo sapien	942	27	100.0	312	2	Q6BPY3_DEBHA	Q6bpy3 debaryomyce
870	27	100.0	288	2	Q9RI48_STRCO	Q9ri48 streptomyce	943	27	100.0	312	2	Q9HWC0_PSEAE	Q9hwc0 pseudomonas
871	27	100.0	289	1	KLF13_MOUSE	Q9ijz6 mus musculus	944	27	100.0	312	2	Q9CNX6_PASMU	Q9cnx6 pasceurella
872	27	100.0	290	2	Q5PSH7_AZOSE	Q5psh7 azoarcus sp	945	27	100.0	313	2	Q9P5K2_NEUCR	Q9p5k2 neurospora
873	27	100.0	290	2	Q9K401_STRCO	Q9k401 streptomyce	946	27	100.0	313	2	Q6YUUS_ORYSA	Q6yuu5 oryza sativ
874	27	100.0	291	2	Q5SNM2_ORYSA	Q5snm2 oryza sativ	947	27	100.0	313	2	Q5UNW3_MIMIV	Q5unw3 mimivirus
875	27	100.0	291	2	Q8GPN6_STRAM	Q8gpn6 streptomyce	948	27	100.0	313	2	Q4ENI4_TETNG	Q4rni4 tetraodon n
876	27	100.0	291	2	Q6WIM6_RHISN	Q6wim6 rhizobium s	949	27	100.0	314	2	Q4UQX8_XANCP	Q4uqx8 xanthomonas
877	27	100.0	292	2	Q5I6A3_PIG	Q5i6a3 sus scrofa	950	27	100.0	314	2	Q4NV06_9DELT	Q4nv06 aeromyxob
878	27	100.0	293	2	Q5VMT6_ORYSA	Q5vmt6 oryza sativ	951	27	100.0	314	2	Q4NV09_9BURK	Q4nv09 burkholderi
879	27	100.0	294	2	Q60MC8_CABER	Q60mc8 caenorhabdi	952	27	100.0	314	2	Q63WA3_XURPS	Q63wa3 burkholderi
880	27	100.0	294	2	Q9BDW9_MACFA	Q9bdw9 macaca fasc	953	27	100.0	314	2	Q8PCJ3_XANCP	Q8pcj3 xanthomonas
881	27	100.0	294	2	Q5Z8B1_ORYSA	Q5z8b1 oryza sativ	954	27	100.0	315	2	Q944A2_ARATH	Q944a2 arabidopsis
882	27	100.0	294	2	Q5SLR8_THET8	Q5slr8 thermus the	955	27	100.0	316	2	Q8S0W7_ORYSA	Q8s0w7 oryza sativ
883	27	100.0	294	2	Q72GT7_THET2	Q72gt7 thermus the	956	27	100.0	316	2	Q5SFC6_STRBI	Q5sfc6 streptomyce
884	27	100.0	295	2	Q52D98_MAGGR	Q52d98 magnaporthe	957	27	100.0	316	2	Q4HC33_9DEIO	Q4hc33 deinococcus
885	27	100.0	295	2	Q20927_CABEL	Q20927 caenorhabdi	958	27	100.0	316	2	Q6OAG9_METCA	Q6oag9 methylococc
886	27	100.0	295	2	Q5VP20_ORYSA	Q5vp20 oryza sativ	959	27	100.0	317	2	Q5ZD27_ORYSA	Q5zd27 oryza sativ
887	27	100.0	295	2	Q84SQ9_ORYSA	Q84sq9 oryza sativ	960	27	100.0	318	2	Q5E226_VIBF1	Q5e226 vibrio fisc
888	27	100.0	295	2	Q7X495_CHLFR	Q7x495 chlorogloe	961	27	100.0	318	2	Q5E226_VIBF1	Q5e226 vibrio fisc
889	27	100.0	295	2	Q7X497_9CYAN	Q7x497 fischerella	962	27	100.0	318	2	Q4HA02_9DEIO	Q4ha02 deinococcus
890	27	100.0	295	2	Q7X498_9NOSO	Q7x498 nostoc ento	963	27	100.0	318	2	Q8PRA9_XANAC	Q8pra9 xanthomonas
891	27	100.0	295	2	Q7X499_9CYAN	Q7x499 scytonema s	964	27	100.0	319	2	Q9VPT9_DROME	Q9vpt9 drosophila
892	27	100.0	295	2	Q7X4A1_9NOST	Q7x4a1 anabaena pl	965	27	100.0	319	2	Q4LTJ1_9BURK	Q4ltj1 burkholderi
893	27	100.0	295	2	Q7X4A2_9NOST	Q7x4a2 synecococc	966	27	100.0	319	2	Q8SQJ1_PSEPK	Q8sqj1 pseudomonas
894	27	100.0	295	2	Q8L131_9CYAN	Q8l131 leptolyngby	967	27	100.0	319	2	Q8EK84_SHEON	Q8ek84 shewanella
895	27	100.0	295	2	Q8L133_9CYAN	Q8l133 pleurocapsa	968	27	100.0	319	2	Q8ZAN5_YERPE	Q8zan5 yersinia pe
896	27	100.0	295	2	Q8L134_NODSP	Q8l134 nodularia s	969	27	100.0	319	2	Q66FR2_YERPS	Q66fr2 yersinia ps
897	27	100.0	295	2	Q8L135_FISMU	Q8l135 fischerella	970	27	100.0	319	2	Q6DAN9_ERWCT	Q6dan9 erwinia car
898	27	100.0	295	2	Q8L138_9CYAN	Q8l138 chlorogloe	971	27	100.0	319	2	Q889Y7_PSESM	Q889y7 pseudomonas
899	27	100.0	295	2	Q8L139_9CYAN	Q8l139 stanieria c	972	27	100.0	319	2	Q7WY66_PHOLL	Q7wye6 photorhabdu
900	27	100.0	295	2	Q8L140_9CYAN	Q8l140 pseudanaba	973	27	100.0	319	2	Q4S168_TETNG	Q4s168 tetraodon n
901	27	100.0	295	2	Q8L141_9CYAN	Q8l141 phormidium	974	27	100.0	320	1	BIRA_SALTY	P37416 salmonella
902	27	100.0	295	2	Q8L142_SYNLE	Q8l142 synecococc	975	27	100.0	320	2	Q5N9H0_ORYSA	Q5n9h0 oryza sativ
903	27	100.0	295	2	Q8L143_9NOST	Q8l143 anabaenopsi	976	27	100.0	320	2	Q7XLS4_ORYSA	Q7xls4 oryza sativ
904	27	100.0	295	2	Q8L144_ANAVA	Q8l144 anabaena va	977	27	100.0	320	2	Q5JPLY_ORYSA	Q5jply oryza sativ
905	27	100.0	295	2	Q8L145_9CHRO	Q8l145 microcystis	978	27	100.0	320	2	Q57H80_SALCH	Q57h80 salmonella
906	27	100.0	295	2	Q8L146_ANACY	Q8l146 anabaena cy	979	27	100.0	320	2	Q4K518_PSEFS	Q4k518 pseudomonas
907	27	100.0	295	2	Q8L147_NOSLI	Q8l147 nostoc linc	980	27	100.0	320	2	Q5PK77_SALPA	Q5pk77 salmonella

```

981 Q61lv3 photobacter
982 Q9kv39 vibrio chol
983 Q8z317 salmonella
984 P06709 escherichia
985 Q4upf8 xanthomonas
986 Q41wt5 azotobacter
987 Q8fb87 escherichia
988 Q83pc6 shigella fl
989 Q87kp6 vibrio para
990 Q8x709 escherichia
991 Q8p3y0 xanthomonas
992 Q8pf94 xanthomonas
993 Q9d0b1 mus musculus
994 Q23796 chironomus
995 Q32511 deinococcus
996 Q73ux1 mycobacteri
997 Q8u8u4 agrobacteri
998 Q7m7i4 vibrio vuln
999 Q55ps4 cryptococcu
1000 Q5kdn2 cryptococcu

```

## ALIGNMENTS

```

RESULT 1
Q6SV46 CHLRE PRELIMINARY; PRT; 17 AA.
AC Q6SV46_
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE O5-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Ribosomal protein S14 (Fragment).
GN Name=RPS14;
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae;
OC Chlamydomonadales; Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CC-2290;
RX MEDLINE=22570934; PubMed=12684385; DOI=10.1128/EC.2.2.362-379.2003;
RA Kathir P., LaVoie M., Brazelton W.J., Haas N.A., Lefebvre P.A.,
RA Silflow C.D.;
RT "Molecular map of the Chlamydomonas reinhardtii nuclear genome.";
RL Eukaryot. Cell 2:362-379(2003).
DR EMBL; AY449707; AAR18091.1; -; Genomic_DNA.
DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
KW Ribosomal protein.
FT NON_TER 1
SQ SEQUENCE 17 AA; 1982 MW; 460839D2269F1C39 CRC64;

```

```

Query Match 100.0%; Score 27; DB 2; Length 17;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GRRGR 5
DB 11 GRRGR 15

```

```

RESULT 2
HSP3 OCTVU STANDARD; PRT; 24 AA.
AC P8315;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 10-MAY-2005 (Rel. 47, Last annotation update)
DE Sperm protamine P3 (Po3) (Fragment).
OS Octopus vulgaris (Octopus).
OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Neocoleoidea;
OC Octopodiformes; Octopoda; Incirrata; Octopodidae; Octopus.
OX NCBI_TaxID=6645;

```

```

[1]
RN RP PROTEIN SEQUENCE, FUNCTION, AND MASS SPECTROMETRY.
RC TISSUE=Sperm;
RX PubMed=15095345; DOI=10.1002/mrd.20068;
RA Gimenez-Bonafe P., Soler F.M., Buesa C., Sautiere P.E., Ausio J.,
RA Kouach M., Kasinsky H.E., Chiwa M.;
RT "Chromatin organization during spermiogenesis in Octopus vulgaris. II:
DNA-interacting proteins.";
RL Mol. Reprod. Dev. 68:232-239(2004).
CC -!- FUNCTION: Protamines substitute for histones in the chromatin of
sperm during the haploid phase of spermatogenesis. They compact
sperm DNA into a highly condensed, stable and inactive complex.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: Testis.
CC -!- MASS SPECTROMETRY: MW=4389; METHOD=Electrospray; RANGE=1-?;
NOTE=Ref.1.
-----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use as long as its content is in no way modified and this statement is not
removed.
-----
CC GO; GO:0000786; C:nucleosome; NAS.
CC GO; GO:0005634; C:nucleus; NAS.
CC GO; GO:0003677; F:DNA binding; NAS.
CC GO; GO:0007001; P:chromosome organization and biogenesis (sen. .; NAS.
CC GO; GO:0007076; P:mitotic chromosome condensation; NAS.
CC GO; GO:0006334; P:nucleosome assembly; NAS.
CC GO; GO:0007283; P:spermatogenesis; NAS.
CC GO; GO:0007283; P:spermatogenesis; NAS.
KW Chromosomal protein; Developmental protein; Differentiation;
KW Direct protein sequencing; DNA condensation; DNA-binding;
KW Nuclear protein; Nucleosome core; Spermatogenesis.
FT COMPIAS 1 16 Poly-Arg.
FT NON_TER 24 24
SQ SEQUENCE 24 AA; 3381 MW; 308E90ED9D2C9C9C CRC64;

```

```

Query Match 100.0%; Score 27; DB 1; Length 24;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GRRGR 5
DB 14 GRRGR 18

```

```

RESULT 3
Q6TQT6 YEAST PRELIMINARY; PRT; 24 AA.
ID Q6TQT6_YEAST PRELIMINARY;
AC Q6TQT6;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE YHR065Cp (Fragment).
GN OrderedLocusNames=YHR065C;
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AB972;
RA Kennedy M.C., Dietrich F.S.;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY389302; AAO97234.1; -; mRNA.
KW Complete proteome.
FT NON_TER 1
SQ SEQUENCE 24 AA; 2866 MW; 83820AB41EF59E7C CRC64;

```

```

Query Match 100.0%; Score 27; DB 2; Length 24;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```





RA Chevallier P., Briand G.;  
RT "A corrected primary structure for dog-fish Scylliorhinus caniculus  
RL protamine 23.";  
RL Biochim. Biophys. Acta 1162:99-104(1993).  
RN [3]  
RP PROTEIN SEQUENCE OF 1-5.  
RX MEDLINE=84000513; PubMed=6615852; DOI=10.1016/0167-4838(83)90031-6;  
RA Guesse M., Sautiere P., Chauviere M., Chevallier P.;  
RT "Extraction, purification and characterization of the sperm protamines  
RL of the dog-fish Scylliorhinus caniculus.";  
RL Biochim. Biophys. Acta 748:93-98(1983).  
CC -!- FUNCTION: Protamines substitute for histones in the chromatin of  
CC sperm during the haploid phase of spermatogenesis. They compact  
CC sperm DNA into a highly condensed, stable and inactive complex.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- TISSUE SPECIFICITY: Testis.  
CC -----  
CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use as long as its content is in no way modified and this statement is not  
CC removed.  
CC -----  
CC PIR: S29829; S29829.  
KW Chromosomal protein; Developmental protein; Differentiation;  
KW Direct protein sequencing; DNA condensation; DNA-binding;  
KW Nuclear protein; Nucleosome core; Spermatogenesis; Testis.  
SQ SEQUENCE 37 AA; 4748 MW; EC2366D6C38D5091 CRC64;  
  
Query Match 100.0%; Score 27; DB 1; Length 37;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02; Indels 0; Gaps 0;  
Matches 5; Conservative 0; Mismatches 0;  
  
QY 1 GRRGR 5  
Db |||||  
30 GRRGR 34  
  
RESULT 7  
Q7LZB0\_9SAUR  
ID Q7LZB0\_9SAUR PRELIMINARY; PRT; 45 AA.  
AC Q7LZB0\_9SAUR PRELIMINARY; PRT; 45 AA.  
DT 01-MAR-2004 (TRENBLrel. 26, Created)  
DT 01-MAR-2004 (TRENBLrel. 26, Last sequence update)  
DE Protamine II-3.  
OS Chrysemys picta.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Testudines; Cryptodira; Testudinoidea; Emydidae; Chrysemys.  
OX NCBI\_TaxID=8479;  
RN [1]  
RP PROTEIN SEQUENCE.  
RX MEDLINE=96394458; PubMed=8798564; DOI=10.1074/jbc.271.38.23547;  
RA Hunt J.G., Kasinsky H.E., Elsey R.M., Wright C.L., Rice P., Bell J.E.,  
RA Sharp D.J., Kiss A.J., Hunt D.F., Arnott D.P., Russ M.M.,  
RA Shabanowitz J., Ausio J.;  
RT "Protamines of reptiles.";  
RL J. Biol. Chem. 271:23547-23557(1996).  
DR PIR: B58208; B58208.  
SQ SEQUENCE 45 AA; 6014 MW; 8C1426F70416CA57 CRC64;  
  
Query Match 100.0%; Score 27; DB 2; Length 45;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GRRGR 5  
Db |||||  
12 GRRGR 16  
  
RESULT 8  
Q7LZB1\_9SAUR  
ID Q7LZB1\_9SAUR PRELIMINARY; PRT; 45 AA.  
AC Q7LZB1\_9SAUR PRELIMINARY; PRT; 45 AA.  
DT 01-MAR-2004 (TRENBLrel. 26, Created)  
DT 01-MAR-2004 (TRENBLrel. 26, Last sequence update)  
DE Protamine II-2.  
OS Chrysemys picta.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Testudines; Cryptodira; Testudinoidea; Emydidae; Chrysemys.  
OX NCBI\_TaxID=8479;  
RN [1]  
RP PROTEIN SEQUENCE.  
RX MEDLINE=96394458; PubMed=8798564; DOI=10.1074/jbc.271.38.23547;  
RA Hunt J.G., Kasinsky H.E., Elsey R.M., Wright C.L., Rice P., Bell J.E.,  
RA Sharp D.J., Kiss A.J., Hunt D.F., Arnott D.P., Russ M.M.,  
RA Shabanowitz J., Ausio J.;  
RT "Protamines of reptiles.";  
RL J. Biol. Chem. 271:23547-23557(1996).  
DR PIR: B58208; B58208.  
SQ SEQUENCE 45 AA; 6014 MW; 8C1426F70416CA57 CRC64;  
  
Query Match 100.0%; Score 27; DB 2; Length 45;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GRRGR 5  
Db |||||  
12 GRRGR 16  
  
RESULT 10  
Q9GT46\_ANOGA  
ID Q9GT46\_ANOGA PRELIMINARY; PRT; 46 AA.  
AC Q9GT46;  
DT 01-MAR-2001 (TRENBLrel. 16, Created)  
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Ribosomal protein S18 (Fragment).  
GN Name=lrps18;  
OS Anopheles gambiae (African malaria mosquito).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae;

AC Q7LZB1;  
DT 01-MAR-2004 (TRENBLrel. 26, Created)  
DT 01-MAR-2004 (TRENBLrel. 26, Last sequence update)  
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
DE Protamine II-2.  
OS Chrysemys picta.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Testudines; Cryptodira; Testudinoidea; Emydidae; Chrysemys.  
OX NCBI\_TaxID=8479;  
RN [1]  
RP PROTEIN SEQUENCE.  
RX MEDLINE=96394458; PubMed=8798564; DOI=10.1074/jbc.271.38.23547;  
RA Hunt J.G., Kasinsky H.E., Elsey R.M., Wright C.L., Rice P., Bell J.E.,  
RA Sharp D.J., Kiss A.J., Hunt D.F., Arnott D.P., Russ M.M.,  
RA Shabanowitz J., Ausio J.;  
RT "Protamines of reptiles.";  
RL J. Biol. Chem. 271:23547-23557(1996).  
DR PIR: B58208; B58208.  
SQ SEQUENCE 45 AA; 5995 MW; 8C04FC970416CA57 CRC64;  
  
Query Match 100.0%; Score 27; DB 2; Length 45;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GRRGR 5  
Db |||||  
12 GRRGR 16  
  
RESULT 9  
Q7LZB5\_9SAUR  
ID Q7LZB5\_9SAUR PRELIMINARY; PRT; 45 AA.  
AC Q7LZB5;  
DT 01-MAR-2004 (TRENBLrel. 26, Created)  
DT 01-MAR-2004 (TRENBLrel. 26, Last sequence update)  
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
DE Protamine II-1.  
OS Chrysemys picta.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Testudines; Cryptodira; Testudinoidea; Emydidae; Chrysemys.  
OX NCBI\_TaxID=8479;  
RN [1]  
RP PROTEIN SEQUENCE.  
RX MEDLINE=96394458; PubMed=8798564; DOI=10.1074/jbc.271.38.23547;  
RA Hunt J.G., Kasinsky H.E., Elsey R.M., Wright C.L., Rice P., Bell J.E.,  
RA Sharp D.J., Kiss A.J., Hunt D.F., Arnott D.P., Russ M.M.,  
RA Shabanowitz J., Ausio J.;  
RT "Protamines of reptiles.";  
RL J. Biol. Chem. 271:23547-23557(1996).  
DR PIR: B58208; B58208.  
SQ SEQUENCE 45 AA; 6154 MW; E76S96F707D9CA46 CRC64;  
  
Query Match 100.0%; Score 27; DB 2; Length 45;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GRRGR 5  
Db |||||  
12 GRRGR 16  
  
RESULT 10  
Q9GT46\_ANOGA  
ID Q9GT46\_ANOGA PRELIMINARY; PRT; 46 AA.  
AC Q9GT46;  
DT 01-MAR-2001 (TRENBLrel. 16, Created)  
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Ribosomal protein S18 (Fragment).  
GN Name=lrps18;  
OS Anopheles gambiae (African malaria mosquito).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae;

```

OC Anophelinae; Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=G3;
RX MEDLINE=20481919; PubMed=11005829; DOI=10.1073/pnas.180060997;
RA Oduol F., Xu J., Niare O., Natarajan R., Vernick K.D.;
RT "Genes identified by an expression screen of the vector mosquito
RT Anopheles gambiae display differential molecular immune response to
RT malaria parasites and bacteria.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:11397-11402(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=G3;
RA Oduol F.O., Xu J., Niare O., Natarajan R., Vernick K.D.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the ribosomal protein S13P family.
DR EMBL: AP283268; AG15373.1; -; mRNA.
DR GO: GO:0005840; C:ribosome; IEA.
DR GO: GO:0003723; F:RNA binding; IEA.
DR GO: GO:0003735; F:structural constituent of ribosome; IEA.
DR GO: GO:0006412; P:protein biosynthesis; IEA.
DR InterPro: IPR001892; Ribosomal_S13.
DR Pfam: PF00416; Ribosomal_S13_1.
DR PROSITE: PS00646; RIBOSOMAL_S13_1; 1.
DR PROSITE: PSS0159; RIBOSOMAL_S13_2; 1.
KW Ribonucleoprotein; Ribosomal protein.
FT NON_TER 1 1
SQ SEQUENCE 46 AA; 5519 MW; 0B91AEB7D54EFA85 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 46;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5
DB 34 GRGR 38

RESULT 11
Q7LZ47_9SAUR
ID Q7LZ47_9SAUR PRELIMINARY; PRT; 47 AA.
AC Q7LZ47;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Protamine II-5.
OS Chrysomys picta.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Testudines; Cryptodira; Testudinoidae; Emydidae; Chrysomys.
OX NCBI_TaxID=8479;
RN [1]
RP PROTEIN SEQUENCE.
RX MEDLINE=96394458; PubMed=8798564; DOI=10.1074/jbc.271.38.23547;
RA Hunt J.G., Kasinsky H.E., Eisey R.M., Wright C.L., Rice P., Bell J.E.,
RA Sharp D.J., Kiss A.J., Hunt D.F., Arnott D.P., Russ M.M.,
RA Shabunowitz J., Ausio J.;
RT "Protamines of reptiles.";
RL J. Biol. Chem. 271:23547-23557(1996).
DR PIR: F58208; F58208.
SQ SEQUENCE 47 AA; 6292 MW; D35C85F586F60416 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 47;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5
DB 12 GRGR 16

RESULT 12
Q9GMB9_RABIT

```

```

ID Q9GMB9_RABIT PRELIMINARY; PRT; 48 AA.
AC Q9GMB9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative collagen XI (Fragment).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Lagomorpha; Leporidae;
OC Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=New Zealand White rabbit; TISSUE=Sclera;
RA Thu M.K., Beuerman R.W.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF295598; AAG03083.1; -; mRNA.
DR GO: GO:0005737; C:cytoplasm; IEA.
DR GO: GO:0005515; F:protein binding; IEA.
DR GO: GO:0007155; P:cell adhesion; IEA.
DR GO: GO:0006817; P:phosphate transport; IEA.
DR InterPro: IPR008161; Clg_helix.
DR InterPro: IPR008160; Collagen.
DR Pfam: PF01391; Collagen_1.
DR ProDom: PD000007; Clg_helix; 1.
KW Collagen.
FT NON_TER 1 1
FT NON_TER 48 48
SQ SEQUENCE 48 AA; 4685 MW; 01C42FD29E7BAB5D CRC64;

Query Match 100.0%; Score 27; DB 2; Length 48;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5
DB 4 GRGR 8

RESULT 13
Q62CY0_BURMA
ID Q62CY0_BURMA PRELIMINARY; PRT; 48 AA.
AC Q62CY0;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein.
GN OrderedLocusNames=BMAA0710;
OS Burkholderia mallei (Pseudomonas mallei).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=13373;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 23344;
RX PubMed=15377793; DOI=10.1073/pnas.0403306101;
RA Nierman W.C., DeShazer D., Kim H.S., Tettelin H., Nelson K.E.,
RA Feldblyum T.V., Ulrich R.L., Ronning C.M., Brinkac L.M.,
RA Daugherty S.C., Davidsen T.D., Deboy R.T., Dimitrov G., Dodson R.J.,
RA Durkin A.S., Gwinn M.D., Haft D.H., Khouri H.M., Kolonay J.F.,
RA Madupu R., Mohammed Y., Nelson W.C., Radune D., Romero C.M.,
RA Sarria S., Selengut J., Shamblin C., Sullivan S.A., White O., Yu Y.,
RA Zafar N., Zhou L., Fraser C.M.;
RT "Structural flexibility in the Burkholderia mallei genome.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:14246-14251(2004).
DR EMBL: CP000011; AAU46875.1; -; Genomic_DNA.
DR TIGR: BMAA0710; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 48 AA; 5131 MW; EEE172D59B4C2D8 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 48;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

SQ SEQUENCE 50 AA; 5889 MW; DE05635A6A3419ED CRC64;

Query Match 100.0%; Score 27; DB 2; Length 50;  
Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
|||||  
DB 36 GRRGR 40

RESULT 16  
Q4GXQ0 9COLE PRELIMINARY; PRT; 51 AA.  
ID Q7LZL4\_XENLA  
Q7LZL4\_XENLA  
AC Q7LZL4\_ PRELIMINARY; PRT; 51 AA.  
DT 01-MAR-2004 (TrEMBLrel. 26, Created)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Ribosomal protein S18, cytosolic (Fragment).  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;  
OC Xenopodinae; Xenopus; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RA Grossberger D., Flajnik M., Marcuz A.;  
RT "Ribosomal and chromosomal protein cDNA clones of Xenopus laevis  
thymus isolated with differential screening."  
RL Comp. Biochem. Physiol. 98:127-133(1991).  
CC -!- SIMILARITY: Belongs to the ribosomal protein S13P family.  
DR PIR; C61510; C61510.  
DR GO; GO:0005840; C:ribosome; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003735; F:structural constituent of ribosome; IEA.  
DR GO; GO:0006412; P:protein biosynthesis; IEA.  
DR InterPro; IPR001892; Ribosomal\_S13.  
DR Pfam; PF00416; Ribosomal\_S13; 1.  
DR ProDom; PD001363; Ribosomal\_S13; 1.  
DR PROSITE; PS0159; RIBOSOMAL\_S13\_2; 1.  
KW Ribonucleoprotein; Ribosomal protein.  
FT NON\_TER 1  
FT NON\_TER 51  
SQ SEQUENCE 51 AA; 5967 MW; 4A1A8287D292B03 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 51;  
Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
|||||  
DB 39 GRRGR 43

RESULT 17  
Q8RYU2\_ORYSA PRELIMINARY; PRT; 53 AA.  
ID Q8RYU2\_ORYSA PRELIMINARY; PRT; 53 AA.  
AC Q8RYU2;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Hypothetical protein B1156H12.5.  
GN Name=B1156H12.5;  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=39947;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RA Sasaki T., Matsumoto T., Yamamoto K., Sakata K., Katayose Y.,  
RA Waki J., Nishimura Y., Cheng Z., Nagamura Y., Antonio B.A., Kanamori H.,  
RA Hosokawa S., Masukawa M., Arikawa K., Chiden Y., Hayashi M.,  
NON\_TER 1

QY 1 GRRGR 5  
|||||  
DB 39 GRRGR 43

RESULT 18  
Q4GXQ0 9COLE PRELIMINARY; PRT; 49 AA.  
ID Q4GXQ0;  
AC Q4GXQ0;  
DT 13-SEP-2005 (TrEMBLrel. 31, Created)  
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)  
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)  
DE Ribosomal protein S18e (Fragment).  
GN Name=rpS18e;  
OS Eucinetus sp. APV-2005.  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Coleoptera; Polyphaga; Elateriformia;  
OC Scitoidae; Eucinetidae; Eucinetus.  
OC NCBI\_TaxID=292446;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RA Longhorn S.J., Vogler A.P.;  
RT "Ribosomal proteins of Coleoptera."  
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AM048973; CAJ7223.1; -, mRNA.  
FT NON\_TER 1  
SQ SEQUENCE 49 AA; 5859 MW; 75E2733548ECC274 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 49;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5  
|||||  
DB 37 GRRGR 41

RESULT 19  
Q4X611\_PLACH PRELIMINARY; PRT; 50 AA.  
ID Q4X611;  
AC Q4X611;  
DT 13-SEP-2005 (TrEMBLrel. 31, Created)  
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)  
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)  
DE Hypothetical protein (Fragment).  
GN ORFNames=PC302493.00.0;  
OS Plasmodium chabaudi.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5825;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RA Hall N., Karras M., Raine J.D., Carlton J.M., Kooij T.W.A.,  
RA Berriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,  
RA James K., Rutherford K., Harris B., Huan H., Church C.,  
RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,  
RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,  
RA Janse C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.;  
RT "A comprehensive survey of the Plasmodium life cycle by genomic,  
transcriptomic, and proteomic analyses."  
RL Science 307:82-86(2005).  
CC -!- CAUTION: The sequence shown here is derived from an  
EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
preliminary data.  
CC -!- SIMILARITY: Belongs to the ribosomal protein S13P family.  
CC EMBL; CAAJ01009502; CAH87495.1; -, Genomic\_DNA.  
DR InterPro; IPR001892; Ribosomal\_S13.  
DR ProDom; PD001363; Ribosomal\_S13; 1.  
DR PROSITE; PS00646; RIBOSOMAL\_S13\_1; 1.  
DR PROSITE; PS0159; RIBOSOMAL\_S13\_2; 1.  
KW Hypothetical protein; Ribonucleoprotein; Ribosomal protein.  
FT NON\_TER 1

```
RA Okamoto M., Ando T., Aoki H., Arita K., Hamada M., Harada C.,
RA Hijishita S., Honda M., Ichikawa Y., Idonuma A., Iijima M., Ikeda M.,
RA Ikono M., Itoh S., Itoh T., Itoh Y., Iwabuchi A., Kamiya K.,
RA Karaawa W., Katagiri S., Kikuta A., Kobayashi N., Kono I.,
RA Machita K., Machara T., Mizuno H., Mizubayashi T., Mukai Y.,
RA Nagasaki K., Nakashima M., Nakama Y., Nakamichi Y., Nakamura M.,
RA Namiki N., Negishi M., Ohta I., Ono N., Saiji S., Sakai K., Shibata M.,
RA Shimokawa T., Shomura A., Song J., Takazaki Y., Terasawa K., Tsuji K.,
RA Waki K., Yamagata H., Yamane H., Yoshiki S., Yoshihara R., Yukawa K.,
RA Zhong H., Iwama H., Endo T., Ito H., Hahn J.H., Kim H.I., Eun M.Y.,
RA Yano M., Jiang J., Gojobori T.,
RT "The genome sequence and structure of rice chromosome 1.";
RL Nature 420:312-316(2002).
DR EMBL; AP004225; BAB90671.1; -; Genomic_DNA.
DR Gramene; Q8RYU2; -.
KW Hypothetical protein.
SQ SEQUENCE 53 AA; 5691 MW; AC911587F53C3A7F CRC64;

Query Match 100.0%; Score 27; DB 2; Length 53;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5
DB 15 GRGR 19

RESULT 18
Q45362_BORPE
ID Q45362_BORPE PRELIMINARY; PRT; 54 AA.
AC Q45362;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Filamentous hemagglutinin antigen (Fragment).
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Mattel D., Pichot F., Bellalou J., Mercereau-Puijalon O., Ullmann A.;
RT "Molecular cloning of a coding sequence of Bordetella pertussis
RT filamentous hemagglutinin gene.";
RL FEMS Microbiol Lett. 37:73-77(1986).
DR EMBL; M35274; AAA22972.1; -; Genomic_DNA.
DR FRL; I40314; I40314. 1
FT NON_TER 54
SQ SEQUENCE 54 AA; 6037 MW; DA3F08D30A538507 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 54;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5
DB 28 GRGR 32

RESULT 19
Q4KK18_PSEF5
ID Q4KK18_PSEF5 PRELIMINARY; PRT; 54 AA.
AC Q4KK18;
DT 13-SEP-2005 (T-EMBLrel. 31, Created)
DT 13-SEP-2005 (T-EMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (T-EMBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORFNames-PFL_0093;
OS Pseudomonas fluorescens (strain Pf-5).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=220664;
```

```
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=pf-5;
RX PubMed=15980861; DOI=10.1038/nbt11110;
RA Paulsen I.T., Press C., Ravel J., Kobayashi D., Myers G.S.,
RA Mavrodii D., DeBoy R.T., Seshadri R., Ren Q., Madupu R., Dodson R.J.,
RA Durkin S., Brinkac L.M., Daugherty S.C., Sullivan S.A., Rosovitz M.,
RA Gwinn M.L., Zhou L., Nelson W.C., Weidman J., Watkins K., Tran K.,
RA Khouri H.M., Pierson E., Pierson L. III, Thomasow L., Loper J.;
RT "Complete genome sequence of the plant commensal Pseudomonas
RT fluorescens Pf-5.";
RL Nat. Biotechnol. 23:873-878(2005).
DR EMBL; CP000076; AAY95510.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 54 AA; 5710 MW; CE9E2C6B5E03331A CRC64;

Query Match 100.0%; Score 27; DB 2; Length 54;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5
DB 49 GRGR 53

RESULT 20
HSPL OCTVU
ID HSPL OCTVU STANDARD; PRT; 56 AA.
AC P832I4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Sperm protamine P1 (P01) [Contains: Sperm protamine P2 (P02) (Main
DE Protamine)].
OS Octopus vulgaris (Octopus).
OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Neocoleoidea;
OC Octopodiformes; Octopoda; Incirrata; Octopodidae; Octopus.
OX NCBI_TaxID=6645;
RN [1]
RP PROTEIN SEQUENCE, FUNCTION, PHOSPHORYLATION, AND MASS SPECTROMETRY.
RC TISSUE=Sperm;
RX PubMed=15095345; DOI=10.1002/mrd.20068;
RA Gimenez-Bonafe P., Soler F.M., Buesa C., Sautiere P.E., Ausio J.,
RA Kouach M., Kasinsky H.E., Chiva M.;
RT "Chromatin organization during spermiogenesis in Octopus vulgaris. II:
RT DNA-interacting proteins.";
RL Mol. Reprod. Dev. 68:232-239(2004).
CC -!- FUNCTION: Protamines substitute for histones in the chromatin of
CC sperm during the haploid phase of spermatogenesis. They compact
CC sperm DNA into a highly condensed, stable and inactive complex.
CC -!- FUNCTION: Octopus spermiogenesis is characterized by a double
CC nuclear protein transition: Histones are first replaced by P1,
CC which allows the chromatin to adopt a shape that is not as relaxed
CC as with histones. The majority of P1 is later replaced by P2,
CC forming a compact chromatin. P2 is the main protamine of sperm.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: Testis.
CC -!- PTM: P2 is phosphorylated in immature sperm. It is dephosphorylated
CC in mature sperm allowing a stronger interaction with DNA.
CC -!- MASS SPECTROMETRY: MW=7428; METHOD=Electrospray; RANGE=1-56;
CC NOTE=Ref.1.
CC -!- MASS SPECTROMETRY: MW=6028; METHOD=Electrospray; RANGE=15-56;
CC NOTE=Ref.1.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC GO; GO:0000786; C:nucleosome; NAS.
CC GO; GO:0005634; C:nucleus; NAS.
CC GO; GO:0003677; F:DNA binding; NAS.
DR
DR
DR
```

```
DR GO: GO:0007001; P:chromosome organization and biogenesis (sen. . .; NAS.
DR GO: GO:0007076; P:mitotic chromosome condensation; NAS.
DR GO: GO:0006334; P:nucleosome assembly; NAS.
DR GO: GO:0007283; P:spermatogenesis; NAS.
KW Chromosomal protein; Developmental protein; Differentiation;
KW Direct protein sequencing; DNA condensation; DNA-binding;
KW Nuclear protein; Nucleosome core; Phosphorylation; Spermatogenesis.
FT CHAIN 1 56 Sperm protamine P1.
FT PROPEP 1 14
FT CHAIN 15 56 Sperm protamine P2.
FT COMPIAS 13 25 Poly-Arg.
FT COMPIAS 35 49 Poly-Arg.
FT COMPIAS 52 56 Poly-Arg.
SQ SEQUENCE 56 AA; 7430 MW; 41EF8A8E024A93D6 CRC64;

Query Match 100.0%; Score 27; DB 1; Length 56;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5
DB 23 GRGR 27

RESULT 21
HSPI_DIDMA STANDARD; PRT; 57 AA.
ID HSPI_DIDMA P35305;
AC P67837;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Sperm protamine P1.
DE Name=PRM1;
OS Didelphis marsupialis virginiana (North American opossum).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Didelphimorphia; Didelphidae; Didelphis.
OX NCBI_TaxID=9267;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Sperm;
RX MEDLINE=93345500; PubMed=8344286;
RA Winkfein R.J., Nishikawa S., Connor W., Dixon G.H.;
RT "Characterization of a marsupial sperm protamine gene and its
RL transcripts from the North American opossum (Didelphis marsupialis).";
RL Eur. J. Biochem. 215:63-72(1993).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95215351; PubMed=7700877;
RA Retief J.D., Krajewski C., Westernman M., Winkfein R.J., Dixon G.H.;
RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
RL Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).
CC -!- FUNCTION: Protamines substitute for histones in the chromatin of
CC sperm during the haploid phase of spermatogenesis. They compact
CC sperm DNA into a highly condensed, stable and inactive complex.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: Testis.
CC -!- SIMILARITY: Belongs to the protamine P1 family.

This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use as long as its content is in no way modified and this statement is not
removed.

EMBL; L17007; AAA02812.1; -; Unassigned DNA.
DR EMBL; X74044; CAA52153.1; -; Genomic_DNA.
DR PIR; S34045; S34045.
DR InterPro; IPR000221; Protamine_P1.
DR Pfam; PF00260; Protamine P1; 1.
DR PROSITE; PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; Developmental protein; Differentiation;
KW DNA condensation; DNA-binding; Nuclear protein; Nucleosome core;
KW Spermatogenesis; Testis.

This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use as long as its content is in no way modified and this statement is not
removed.

EMBL; L17007; AAA02812.1; -; Unassigned DNA.
DR EMBL; X74044; CAA52153.1; -; Genomic_DNA.
DR PIR; S34045; S34045.
DR InterPro; IPR000221; Protamine_P1.
DR Pfam; PF00260; Protamine P1; 1.
DR PROSITE; PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; Developmental protein; Differentiation;
KW DNA condensation; DNA-binding; Nuclear protein; Nucleosome core;
KW Spermatogenesis; Testis.

FT INIT MET 0 0 By similarity.
SQ SEQUENCE 57 AA; 7810 MW; 283715B280214E52 CRC64;

Query Match 100.0%; Score 27; DB 1; Length 57;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5
DB 37 GRGR 41

RESULT 22
HSPI_MONDO STANDARD; PRT; 57 AA.
ID HSPI_MONDO P35305;
AC P67836;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Sperm protamine P1.
DE Name=PRM1;
OS Monodelphis domestica (Short-tailed gray opossum).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Didelphimorphia; Didelphidae; Monodelphis.
OX NCBI_TaxID=13616;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95215351; PubMed=7700877;
RA Retief J.D., Krajewski C., Westernman M., Winkfein R.J., Dixon G.H.;
RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
RL Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).
CC -!- FUNCTION: Protamines substitute for histones in the chromatin of
CC sperm during the haploid phase of spermatogenesis. They compact
CC sperm DNA into a highly condensed, stable and inactive complex.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: Testis.
CC -!- SIMILARITY: Belongs to the protamine P1 family.

This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use as long as its content is in no way modified and this statement is not
removed.

EMBL; L35448; AAA74612.1; -; Genomic_DNA.
DR InterPro; IPR000221; Protamine P1.
DR Pfam; PF00260; Protamine P1; 1.
DR PROSITE; PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; Developmental protein; Differentiation;
KW DNA condensation; DNA-binding; Nuclear protein; Nucleosome core;
KW Spermatogenesis; Testis.
FT INIT MET 0 0 By similarity.
SQ SEQUENCE 57 AA; 7810 MW; 283715B280214E52 CRC64;

Query Match 100.0%; Score 27; DB 1; Length 57;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5
DB 37 GRGR 41

RESULT 23
HSPI_CHRPI STANDARD; PRT; 58 AA.
ID HSPI_CHRPI Q71ZB2;
AC Q71ZB2;
DT 05-JUL-2004 (Rel. 44, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Sperm protamine P1-type (Protamine I-1).
DE Chromosomal protein; Developmental protein; Differentiation;
OS Chrysemys picta bellii (Western painted turtle).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

OC Testudines; Cryptodira; Testudinoidea; Emydidae; Chrysemys.  
OX NCBI\_TaxID=8478;  
RN [1]  
RP PROTEIN SEQUENCE.  
RC TISSUE=Spem;  
RX PubMed=8798564; DOI=10.1074/jbc.271.38.23547;  
RA Hunt J.G., Kasinsky H.E., Elsey R.M., Wright C.L., Rice P., Bell J.E.,  
RA Sharp D.J., Kiss A.J., Hunt D.F., Arnott D.P., Russ M.M.,  
RA Shabanowitz J., Ausio J.;  
RT "Proteins of reptiles";  
RL J. Biol. Chem. 271:23547-23557(1996).  
CC -!- FUNCTION: Proteins substitute for histones in the chromatin of  
CC sperm during the haploid phase of spermatogenesis. They compact  
CC sperm DNA into a highly condensed, stable and inactive complex (By  
CC similarity).  
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
CC -!- TISSUE SPECIFICITY: Testis.  
CC -!- SIMILARITY: Belongs to the protamine P1 family.  
CC -----  
CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use as long as its content is in no way modified and this statement is not  
CC removed.  
CC -----  
CC PIR; A58208; A58208.  
CC InterPro; IPR000221; Protamine P1.  
CC PROSITE; PS00048; PROTAMINE P1; 1.  
CC Chromosomal protein; Developmental protein; Differentiation;  
CC Direct protein sequencing; DNA condensation; DNA-binding;  
CC Nuclear protein; Nucleosome core; Spermatogenesis; Testis.  
SQ SEQUENCE 58 AA; 7620 MW; 3826F4875A692424 CRC64;  
  
Query Match 100.0%; Score 27; DB 1; Length 58;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GRRGR 5  
DB 24 GRRGR 28  
  
RESULT 24  
Q981D3 SULSO  
ID Q981D3 SULSO PRELIMINARY; PRT; 59 AA.  
AC Q981D3;  
DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Hypothetical protein.  
GN OrderedLocusNames=SS05027;  
OS Sulfolobus solfataricus.  
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
OC Sulfolobus.  
OX NCBI\_TaxID=2287;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=ATCC 35092 / DSM 1617 / P2;  
RX MEDLINE=21332296; PubMed=11427726; DOI=10.1073/pnas.141222098;  
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,  
RA Aweyaz M.J., Chan-Weher C.C.-Y., Clausen I.G., Curtis B.A.,  
RA De Moors A., Brauso G., Fletcher C., Gordon P.M.K.,  
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,  
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,  
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,  
RA Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.;  
RA "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).  
DR EMBL; A806643; AKA40379.1; -; Genomic\_DNA.  
DR PIR; D90140; D90140.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 59 AA; 7227 MW; 8488B3A72AC00786 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 59;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GRRGR 5  
DB 48 GRRGR 52  
  
RESULT 25  
Q7QZW6 GIALA  
ID Q7QZW6 GIALA PRELIMINARY; PRT; 60 AA.  
AC Q7QZW6;  
DT 01-MAR-2004 (TrEMBLrel. 26, Created)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)  
DE GLP 609 178 360.  
OS Giardia lamblia ATCC 50803.  
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.  
OX NCBI\_TaxID=184922;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=WB C6;  
RA Morrison H.G., McArthur A.G., Adam R.D., Aley S.B., Gillin F.D.,  
RA Olsen G.J., Sogin M.L.;  
RT "Draft sequence of the Giardia lamblia genome.";  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
DR EMBL; AACB0100040; EAA40549.1; -; Genomic DNA.  
SQ SEQUENCE 60 AA; 6430 MW; 6F024DE6853F2A74 CRC64;  
  
Query Match 100.0%; Score 27; DB 2; Length 60;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GRRGR 5  
DB 6 GRRGR 10  
  
RESULT 26  
Q4T4G0 TETNG  
ID Q4T4G0 TETNG PRELIMINARY; PRT; 63 AA.  
AC Q4T4G0;  
DT 13-SEP-2005 (TrEMBLrel. 31, Created)  
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)  
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)  
DE Chromosome undetermined SCAF9668, whole genome shotgun sequence.  
DE (Fragment).  
GN ORFNames=GSTENG0007340001;  
OS Tetraodon nigroviridis (Green puffer).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
OC Tetraodontidae; Tetraodontidae; Tetraodon.  
OX NCBI\_TaxID=99883;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=O. Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,  
RA Jaillon O., Aury J.M., Brunet F., Fischer C., Ozouf-Costaz C., Bernot A.,  
RA Mauceli E., Bouteau L., Fischer S., Lutfalla G., Dossat C., Segurens B.,  
RA Nicaud S., Jaife D., Fisher S., Lutfalla G., Dossat C., Segurens B.,  
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,  
RA Anthouard V., Jubin C., Castellani V., Katinka M., Vacherie B.,  
RA Biemont C., Skalli Z., Cattolico L., Poulin J., De Berardinis V.,  
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,  
RA Parra G., Lardier G., Chapelle C., McKernan K.J., McEwan P., Bosak S.,  
RA Kellis M., Wolff J.N., Guigo R., Zody M.C., Mesirov J.,  
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,  
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,  
RA Winkler P., Lander E.S., Weissbach J., Roest Crolius H.;  
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals

```
RT the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBSJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBSJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CAAB01009668; CAP92222.1; -; Genomic_DNA.
FT NON TER 1
SQ SEQUENCE 63 AA; 6976 MW; 51DC606082B1E576 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 63;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5
DB 37 GRRGR 41

RESULT 27
Q6ZCY1 ORYSA
ID Q6ZCY1 ORYSA PRELIMINARY; PRT; 64 AA.
AC Q6ZCY1;
DT 05-JUL-2004 (TREMELrel. 27, Created)
DT 05-JUL-2004 (TREMELrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMELrel. 27, Last annotation update)
DE Hypothetical protein P0026A08.20.
GN Name=P0026A08.20;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AP004556; BAC99524.1; -; Genomic_DNA.
DR Gramene; Q6ZCY1; -.
KW Hypothetical protein.
SQ SEQUENCE 64 AA; 7211 MW; EC0B17ED04609CF1 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 64;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5
DB 57 GRRGR 61

RESULT 28
Q4RQE3 TETNG
ID Q4RQE3 TETNG PRELIMINARY; PRT; 64 AA.
AC Q4RQE3;
DT 13-SEP-2005 (TREMELrel. 31, Created)
DT 13-SEP-2005 (TREMELrel. 31, Last sequence update)
DT 13-SEP-2005 (TREMELrel. 31, Last annotation update)
DE Chromosome 17 SCAP15006, whole genome shotgun sequence.
ORFNames=GSTENG00030646001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neoceratgii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
```

```
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Doseat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castellani V., Katinka M., Vacherie B.,
RA Biemont C., Skalli Z., Cattolico L., Poullain J., De Berardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudat V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissbach J., Roest Crolius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBSJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBSJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CAAB01015006; CAG09389.1; -; Genomic_DNA.
SQ SEQUENCE 64 AA; 7104 MW; 3F55BDC473580F7A CRC64;

Query Match 100.0%; Score 27; DB 2; Length 64;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5
DB 58 GRRGR 62

RESULT 29
RS14 TOBAC
ID RS14 TOBAC STANDARD; PRT; 66 AA.
AC P93377;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE 40S ribosomal protein S14 (Fragment).
GN Name=RPS14;
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
OC asterids; lamids; Solanales; Solanales; Solanales; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Karrer E.E., Beachy R.N., Holt C.A.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBSJ databases.
CC -!- SIMILARITY: Belongs to the ribosomal protein S1P family.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
DR EMBL; U62622; AAC49968.1; -; mRNA.
DR SMR; P93377; 1-62.
DR InterPro; IPR001971; Ribosomal_S11.
DR PANTHER; PTHR11759; Ribosomal_S11; 1.
DR Pfam; PF00411; Ribosomal_S11; 1.
DR ProDom; PD001010; Ribosomal_S11; 1.
DR PROSITE; PS00054; RIBOSOMAL_S11; 1.
KW Ribonucleoprotein; Ribosomal protein.
FT NON TER 1
SQ SEQUENCE 66 AA; 7000 MW; B94E9C68FBDBAC23 CRC64;

Query Match 100.0%; Score 27; DB 1; Length 66;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```



```

QY      1 GRRGR 5
Db      60 GRRGR 64

RESULT 30
Q6YWP2 ORYSA
ID Q6YWP2 ORYSA PRELIMINARY; PRT; 66 AA.
AC Q6YWP2
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 01-FEB-2005 (TREMBlrel. 29, Last annotation update)
DE Hypothetical protein OSJNBa0089L03.14 (Hypothetical protein
DE OSJNBa0003D23.32).
DE OSJNBa0003D23.32).
GN Name=OSJNBa0089L03.14; Synonyms=OSJNBa0003D23.32;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Katayose Y.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005754; BAD13257.1; -; Genomic DNA.
DR EMBL; AP005488; BAD11621.1; -; Genomic DNA.
DR Gramene; O6YWP2; -.
KW Hypothetical protein.
SQ SEQUENCE 66 AA; 7007 MW; B95229719613F48D CRC64;

Query Match 100.0%; Score 27; DB 2; Length 66;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GRRGR 5
Db      11 GRRGR 15

RESULT 31
Q85SM9 ORYSA
ID Q85SM9 ORYSA PRELIMINARY; PRT; 68 AA.
AC Q85SM9;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-FEB-2005 (TREMBlrel. 29, Last annotation update)
DE Hypothetical protein OJ1003C07.10.
GN ORFNAMES=OJ1003C07.10;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA McCombie W.R., Spiegel L., de la Bastide M., Preston R., Ferraro K.,
RA Kuit K., Nascimben L., Zutavern T., Balija V., Bell M., Baker J.,
RA Santos L., Miller B., Katzenberger F., Muller S., King L., Yang C.,
RA Dike S., O'Shaughnessy A., Palmer L., Dedhia N.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA The Rice Chromosome 10 Sequencing Consortium;
RA "In-depth view of structure, activity, and evolution of rice
RT chromosome 10.";
RL Science 300:1566-1569(2003).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RA Buell C.R., Wing R.A., McCombie W.R., Messing J., Yuan Q.;

```

---

```

RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC113335; AM08830.1; -; Genomic DNA.
DR EMBL; AE010884; AAP53370.1; -; Genomic DNA.
DR Gramene; Q85SM9; -.
KW Hypothetical protein.
SQ SEQUENCE 68 AA; 6515 MW; E414A59656C96F6 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 68;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GRRGR 5
Db      50 GRRGR 54

RESULT 32
Q9FRQ5 ORYSA
ID Q9FRQ5 ORYSA PRELIMINARY; PRT; 70 AA.
AC Q9FRQ5;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Putative cellulose synthase 5-partial (Fragment).
GN Name=OSJNBa0056G17.16;
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Burr P.C., Haiao J.,
RA Zismann V., Pai G., Bowman C.L., Fujii C.Y., VanAken S.E.,
RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblyum T.V.,
RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Buell R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC018727; AAG46166.1; -; Genomic DNA.
DR Gramene; Q9FRQ5; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0016760; F:cellulose synthase (UDP-forming) activity; IEA.
DR GO; GO:0030244; P:cellulose biosynthesis; IEA.
DR InterPro; IPR005150; Cellulose synth.
DR Pfam; PF03552; Cellulose_synt; 1.
FT NON_TER 1
SQ SEQUENCE 70 AA; 7639 MW; 1A32CA45E600C222 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 70;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GRRGR 5
Db      25 GRRGR 29

RESULT 33
Q9QVW0 9MURI
ID Q9QVW0 9MURI PRELIMINARY; PRT; 70 AA.
AC Q9QVW0;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE HOXB-6=HOMBOBOX protein (fragment).
OS Mus sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10095;

```

(1)
RN NUCLEOTIDE SEQUENCE.
RP MEDLINE=95042335; PubMed=7954431;
RA Friedmann Y., Daniel C.A., Strickland P., Daniel C.W.;
RT "Hox genes in normal and neoplastic mouse mammary gland.";
RL Cancer Res. 54:5981-5985(1994).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
DR HSP; P02833; 1HOM.
DR SMR; Q9QVV0; 21-70.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001827; Antennapedia.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR012287; Homeodomain-rel.
DR Pfam; PF00046; Homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00032; ANTENNAPEIDIA; UNKNOWN\_1.
DR PROSITE; PS0071; HOMEBOX\_2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
FT NON\_TER 1 1
FT NON\_TER 70 70
SQ SEQUENCE 70 AA; 8706 MW; 57FB96A6AB5C20A7 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 70;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5
DB 20 GRRGR 24

RESULT 34
Q9PRL3\_9CALL PRELIMINARY; PRT; 70 AA.
AC Q9PRL3;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2004 (TREMBlrel. 26, Last annotation update)
DE Homeobox gene product (Fragment).
OS Gallus sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI\_TaxID=9036;
RN [1]\_TaxID=9036;
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=92037185; PubMed=1682126;
RA Mackem S., Mahon K.A.;
RT "Glox 4.7: a chick homeobox gene expressed primarily in limb buds with
limb-type differences in expression.";
RL Development 112:791-806(1991).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
DR HSP; P02833; 1HOM.
DR SMR; Q9PRL3; 21-70.
DR Ensemble; ENSGALG00000011059; Gallus gallus.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001827; Antennapedia.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR012287; Homeodomain-rel.
DR Pfam; PF00046; Homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00032; ANTENNAPEIDIA; UNKNOWN\_1.
DR PROSITE; PS0071; HOMEBOX\_2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
FT NON\_TER 1 1

FT NON\_TER 70 70
SQ SEQUENCE 70 AA; 8642 MW; 2FD797BCD832B4F9 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 70;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5
DB 20 GRRGR 24

RESULT 35
Q4TTN3\_STRM PRELIMINARY; PRT; 71 AA.
AC Q4TTN3;
DT 13-SEP-2005 (TREMBlrel. 31, Created)
DT 13-SEP-2005 (TREMBlrel. 31, Last sequence update)
DT 13-SEP-2005 (TREMBlrel. 31, Last annotation update)
DE Putative RNA polymerase beta prime (Fragment).
OS Streptomyces rimosus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI\_TaxID=1927;
RN [1]\_TaxID=1927;
RP NUCLEOTIDE SEQUENCE.
RA Hejazi M.S., Zeresnki Nobar L., Azarbaijani R.,
RA Mohammadzadeh Sadigh Y.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; DQ062982; AAY58899.1; -; Genomic\_DNA.
FT NON\_TER 1 1
FT NON\_TER 71 71
SQ SEQUENCE 71 AA; 7934 MW; 290B61C5BBS39925 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 71;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5
DB 24 GRRGR 28

RESULT 36
Q6K3X4\_ORYSA PRELIMINARY; PRT; 73 AA.
AC Q6K3X4;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Hypothetical protein P0692F07.9.
GN Name=P0692F07.9;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI\_TaxID=39947;
RN [1]\_TaxID=39947;
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Hattori M., Sakaki Y., Katayose Y.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005593; BAD23446.1; -; Genomic\_DNA.
DR Gramene; Q6K3X4; -.
KW Hypothetical protein.
SQ SEQUENCE 73 AA; 8421 MW; C4DF8A61A848384C CRC64;

Query Match 100.0%; Score 27; DB 2; Length 73;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRRGR 5
DB 47 GRRGR 51

```

Db      67 GRRGR 71

RESULT 37
Q69VE1_ORYSA PRELIMINARY;      PRT;      74 AA.
AC Q69VE1_28 (T-EMBLrel. 28, Created)
DT 25-OCT-2004 (T-EMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (T-EMBLrel. 28, Last annotation update)
DE Hypothetical protein P0417D05.16.
GN Name=P0417D05.16;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, PAC
RT clone: P0417D05.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP004236; BAD33040.1; -; Genomic_DNA.
DR Gramene; Q69VE1; -.
KW Hypothetical protein.
SQ SEQUENCE 74 AA; 7459 MW; 2E72A907BBC59B46 CRC64;

Query Match      100.0%; Score 27; DB 2; Length 74;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GRRGR 5
Db      25 GRRGR 29

RESULT 38
Q8SA4_ORYSA PRELIMINARY;      PRT;      74 AA.
AC Q8SA4_21 (T-EMBLrel. 21, Created)
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Hypothetical protein OSUNBa0010R04.18.
GN Name=OSUNBa0010R04.18;
OS Oryza sativa (rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Buell C.R., Yuan Q., Ouyang S., Liu J., Moffat K.S., Hill J.N.,
RA Gansberger K., Brenner M., Burgess S., Hance M., Shvartsbeyn M.,
RA Tsitrin T., Riggs F., Hsiao J., Ziemann V., Blunt S., Pai G.,
RA VanAken S.E., Utterback T.R., Feldblyum T.V., Kalb E., Quackenbush J.,
RA Salzberg S.L., White O., Fraser C.M.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Buell R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC096687; AAL79737.1; -; Genomic_DNA.
DR Gramene; Q8SA4; -.
KW Hypothetical protein.
SQ SEQUENCE 74 AA; 7187 MW; F3638569F3CA50C1 CRC64;

Query Match      100.0%; Score 27; DB 2; Length 74;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GRRGR 5
Db      25 GRRGR 29

RESULT 39
Q54391_STRLI PRELIMINARY;      PRT;      75 AA.
AC Q54391_01 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE RecA gene. (Fragment).
OS Streptomyces lividans.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1916;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN-TK24;
RX MEDLINE=94283869; PubMed=8013883; DOI=10.1016/0378-1097(94)90596-7;
RA Nussebaumer B., Wohlleben W.;
RT "Identification, isolation and sequencing of the recA gene of
RT Streptomyces lividans TK24.";
RL FEMS Microbiol. Lett. 118:57-63(1994).
DR EMBL: X76076; CAAS3675.1; -; Genomic_DNA.
DR FIR; S38675; S38675. 1
FT NON TER
SQ SEQUENCE 75 AA; 8176 MW; 0A3BFF0C253819FB CRC64;

Query Match      100.0%; Score 27; DB 2; Length 75;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GRRGR 5
Db      22 GRRGR 26

RESULT 40
Q6ZFA5_ORYSA PRELIMINARY;      PRT;      76 AA.
AC Q6ZFA5_27 (T-EMBLrel. 27, Created)
DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)
DE Hypothetical protein OJ1224_G08.13.
GN Name=OJ1224_G08.13;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 8, BAC
RT clone: OJ1224_G08.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP004256; BAD01682.1; -; Genomic_DNA.
DR Gramene; Q6ZFA5; -.
KW Hypothetical protein.
SQ SEQUENCE 76 AA; 8574 MW; F1AB01E32294E5 CRC64;

Query Match      100.0%; Score 27; DB 2; Length 76;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GRRGR 5
Db      26 GRRGR 30

RESULT 41
Q8FS22_COREF
```

```

Db      67 GRRGR 71

RESULT 39
Q54391_STRLI PRELIMINARY;      PRT;      75 AA.
AC Q54391_01 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE RecA gene. (Fragment).
OS Streptomyces lividans.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1916;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN-TK24;
RX MEDLINE=94283869; PubMed=8013883; DOI=10.1016/0378-1097(94)90596-7;
RA Nussebaumer B., Wohlleben W.;
RT "Identification, isolation and sequencing of the recA gene of
RT Streptomyces lividans TK24.";
RL FEMS Microbiol. Lett. 118:57-63(1994).
DR EMBL: X76076; CAAS3675.1; -; Genomic_DNA.
DR FIR; S38675; S38675. 1
FT NON TER
SQ SEQUENCE 75 AA; 8176 MW; 0A3BFF0C253819FB CRC64;

Query Match      100.0%; Score 27; DB 2; Length 75;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GRRGR 5
Db      22 GRRGR 26

RESULT 40
Q6ZFA5_ORYSA PRELIMINARY;      PRT;      76 AA.
AC Q6ZFA5_27 (T-EMBLrel. 27, Created)
DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)
DE Hypothetical protein OJ1224_G08.13.
GN Name=OJ1224_G08.13;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 8, BAC
RT clone: OJ1224_G08.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP004256; BAD01682.1; -; Genomic_DNA.
DR Gramene; Q6ZFA5; -.
KW Hypothetical protein.
SQ SEQUENCE 76 AA; 8574 MW; F1AB01E32294E5 CRC64;

Query Match      100.0%; Score 27; DB 2; Length 76;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GRRGR 5
Db      26 GRRGR 30

RESULT 41
Q8FS22_COREF
```

ID	Q8FS22_COREF	PRELIMINARY;	PRT;	76 AA.
AC	Q8FS22;			
DT	01-MAR-2003	(TREMBLrel. 23, Created)		
DT	01-MAR-2003	(TREMBLrel. 23, Last sequence update)		
DT	01-MAR-2003	(TREMBLrel. 23, Last annotation update)		
DE	Hypothetical protein.			
GN	OrderedLocusNames=CS0584;			
OS	Corynebacterium efficiens.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacterineae; Corynebacteriaceae; Corynebacterium.			
OX	NCBI_TaxID=152794;			
RN	[1]			
RP	NUCLEOTIDE SEQUENCE.			
RC	STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;			
RX	MEDLINE=22723752; PubMed=12840036; DOI=10.1101/gr.1285603;			
RA	Niehio Y., Nakamura Y., Kawarabayashi Y., Usuda Y., Kimura E.,			
RA	Suginoto S., Matsui K., Yanagishi A., Kikuchi H., Ikeo K.,			
RA	Gotohori T.;			
RT	"Comparative complete genome sequence analysis of the amino acid			
RT	replacements responsible for the thermostability of Corynebacterium			
RT	efficiens.";			
RL	Genome Res. 13:1572-1579(2003).			
DR	EMBL; BA000035; BAC17394.1; -; Genomic DNA.			
KW	Complete proteome; Hypothetical protein.			
SQ	SEQUENCE 76 AA; 7824 MW; 820DB0050ESD5D9D CRC64;			
	Query Match 100.0%; Score 27; DB 2; Length 76;			
	Best Local Similarity 100.0%; Pred. No. 4.2e+02;			
	Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1 GRRGR 5			
Db	61 GRRGR 65			
RESULT 42				
ID	Q67PE1_SYMTH	PRELIMINARY;	PRT;	77 AA.
AC	Q67PE1;			
DT	25-OCT-2004	(TREMBLrel. 28, Created)		
DT	25-OCT-2004	(TREMBLrel. 28, Last sequence update)		
DT	25-OCT-2004	(TREMBLrel. 28, Last annotation update)		
DE	Hypothetical protein.			
GN	OrderedLocusNames=STH1467;			
OS	Symbiobacterium thermophilum.			
OC	Bacteria; Actinobacteria; Symbiobacterium.			
OX	NCBI_TaxID=2734;			
RN	[1]			
RP	NUCLEOTIDE SEQUENCE.			
RC	STRAIN=IAM14863;			
RX	PubMed=15383646; DOI=10.1093/nar/gkh830;			
RA	Ueda K., Yamaehita A., Ishikawa J., Shimada M., Watsuji T.,			
RA	Morimura K., Ikeda H., Hattori M., Beppu T.;			
RT	"Genome sequence of Symbiobacterium thermophilum, an uncultivable			
RT	bacterium that depends on microbial commensalism.";			
RL	Nucleic Acids Res. 32:4937-4944(2004).			
DR	EMBL; AP006840; BAD40452.1; -; Genomic DNA.			
DR	GO; GO:0003676; F.nucleic acid binding; IEA.			
DR	InterPro; IPR004088; KH_type_1.			
DR	PROSITE; PS0084; KH_TYPE_1; 1.			
KW	Complete proteome; Hypothetical protein.			
SQ	SEQUENCE 77 AA; 8517 MW; 998E3C967BD6B75A CRC64;			
	Query Match 100.0%; Score 27; DB 2; Length 77;			
	Best Local Similarity 100.0%; Pred. No. 4.2e+02;			
	Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1 GRRGR 5			
Db	47 GRRGR 51			
RESULT 43				
ID	Q7SHB4_NEUCR	PRELIMINARY;	PRT;	79 AA.
AC	Q7SHB4;			
DT	01-MAR-2004	(TREMBLrel. 26, Created)		
DT	01-MAR-2004	(TREMBLrel. 26, Last sequence update)		
DT	01-MAR-2004	(TREMBLrel. 26, Last annotation update)		
DE	Predicted protein.			
GN	Name=NCU01892.1;			
OS	Neurospora crassa.			
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;			
OC	Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.			
OX	NCBI_TaxID=5141;			
RN	[1]			
RP	NUCLEOTIDE SEQUENCE.			
RC	STRAIN=OR74A;			
RA	Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,			
RA	Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,			
RA	Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,			
RA	Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M., Schulte U.,			
RA	Selitrennikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Greenberg D.,			
RA	Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Gnerre S.,			
RA	Roy A., Foley K., Naylor J., Thomann N., Barrett R., Rudd S., Frishman D.,			
RA	Kamal M., Kamvaselis M., Mauceli E., Belke C., Perkins D.D., Kroken S.,			
RA	Krystofova S., Rasmussen C., Metzberg R.L., Perkins D.D., Kroken S.,			
RA	Cogoni C., Macino G., Catchside D., Li W., Pratt R.J., Osmani S.A.,			
RA	Desouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,			
RA	Natvig D.O., Piamann M., Seifer S., Dunlap J., Radford A., Aramayo R.,			
RA	Paulsen I., Sachs M.S., Lander E.S., Nussbaum C., Birren B.;			
RT	"The Genome Sequence of the Filamentous Fungus Neurospora crassa.";			
RL	Nature 0:0-0(2003).			
CC	-!- CAUTION: The sequence shown here is derived from an			
CC	EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is			
CC	preliminary data.			
DR	EMBL; AABX01000004; EAA36287.1; -; Genomic DNA.			
DR	GO; GO:0006412; P:protein biosynthesis; IEA.			
SQ	SEQUENCE 79 AA; 8862 MW; 247492EADBA58326 CRC64;			
	Query Match 100.0%; Score 27; DB 2; Length 79;			
	Best Local Similarity 100.0%; Pred. No. 4.3e+02;			
	Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1 GRRGR 5			
Db	48 GRRGR 52			
RESULT 44				
ID	Q9XSU9_CANFA	PRELIMINARY;	PRT;	79 AA.
AC	Q9XSU9;			
DT	01-NOV-1999	(TREMBLrel. 12, Created)		
DT	01-NOV-1999	(TREMBLrel. 12, Last sequence update)		
DT	01-JUN-2003	(TREMBLrel. 24, Last annotation update)		
DE	Ribosomal protein S14 (Fragment).			
GN	Name=rps14;			
OS	Canis familiaris (Dog).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;			
OC	Canis.			
OX	NCBI_TaxID=9615;			
RN	[1]			
RP	NUCLEOTIDE SEQUENCE.			
RC	TISSUE=Thyroid;			
RX	MEDLINE=20422104; PubMed=10964405; DOI=10.1006/abio.2000.4674;			
RA	Pichon B., Mercan D., Pouillon V., Christophe-Hobertus C.,			
RA	Christophe D.;			
RT	"A method for the large-scale cloning of nuclear proteins and nuclear			
RT	targeting sequences on a functional basis.";			
RL	Anal. Biochem. 284:231-239(2000).			
CC	-!- SIMILARITY: Belongs to the ribosomal protein S11p family.			
DR	EMBL; AJ38514; CAB46816.1; -; mRNA.			
DR	SMR; Q9XSU9; 1-75.			

```
DR Ensembl; ENSCARG00000018094; Canis familiaris.
DR GO; GO:0005840; C:ribosome; IEA.
DR GO; GO:0003735; F:protein biosynthesis; IEA.
DR GO; GO:0006412; P:protein biosynthesis; IEA.
DR InterPro; IPR001971; Ribosomal S11.
DR Pfam; PF00411; Ribosomal S11; 1.
DR ProDom; PD001010; Ribosomal S11; 1.
DR PROSITE; PS00054; RIBOSOMAL_S11; 1.
KW Ribonucleoprotein; Ribosomal protein.
FT NON TER 1
SQ SEQUENCE 79 AA; 8390 MW; 85F9C21B4E7D4D6E CRC64;

Query Match 100.0%; Score 27; DB 2; Length 79;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5
DB 73 GRGR 77

RESULT 45
Q6ERZ5_ORYSA PRELIMINARY; PRT; 79 AA.
AC Q6ERZ5;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein P0698G06.24;
GN Names=P0698G06.24;
OS Oryza sativa (Japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzoaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Katayose Y.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR ENBL; AP005400; BAD28575.1; -; Genomic_DNA.
DR Gramene; Q6ERZ5; -.
KW Hypothetical protein.
SQ SEQUENCE 79 AA; 8690 MW; EDD2E59EED9EC8F CRC64;

Query Match 100.0%; Score 27; DB 2; Length 79;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5
DB 8 GRGR 12

RESULT 46
Q5YS36_NOCFA PRELIMINARY; PRT; 80 AA.
AC Q5YS36;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein.
GN OrderedLocusNames=ncfa41560;
OS Nocardia farcinica.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Nocardiaceae; Nocardia.
OX NCBI_TaxID=37329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=IPM 10152;
RX PubMed=15466710; DOI=10.1073/pnas.0406410101;
RA Ishikawa J., Yamashita A., Mikami Y., Hoshino Y., Kurita H., Hotta K.,
RA Shiba T., Hattori M.;
RT "The complete genomic sequence of Nocardia farcinica IPM 10152.";

Proc. Natl. Acad. Sci. U.S.A. 101:14925-14930 (2004).
DR ENBL; AF006618; BAD59005.1; -; Genomic_DNA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR004088; KH_type_1.
DR Pfam; PF00013; KH_1; 1.
DR PROSITE; PS50084; KH_TYPE_1; 1.
DR PROSITE; PS50084; KH_TYPE_1; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 80 AA; 8433 MW; 13D852405C9793AE CRC64;

Query Match 100.0%; Score 27; DB 2; Length 80;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5
DB 30 GRGR 34

RESULT 47
Q9YMW8_NPVLD PRELIMINARY; PRT; 80 AA.
AC Q9YMW8;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE LdOrf-6 peptide.
OS Lymantria dispar multicapsid nuclear polyhedrosis virus (LdMNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=10449;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=99124785; PubMed=9887315; DOI=10.1006/viro.1998.9469;
RA Kuzio J., Pearson M.N., Harwood S.H., Funk C.J., Evans J.T.,
RA Slavicek J.M., Rohrmann G.F.;
RT "Sequence and analysis of the genome of a baculovirus pathogenic for
LYMANTRIA DISPAR.";
RL Virology 253:17-34 (1999).
DR ENBL; AF081810; AAC70191.1; -; Genomic_DNA.
DR PIR; T30353; T30353.
SQ SEQUENCE 80 AA; 8960 MW; 36B339070492ED3B CRC64;

Query Match 100.0%; Score 27; DB 2; Length 80;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5
DB 6 GRGR 10

RESULT 48
Q73ZQ7_MYCPA PRELIMINARY; PRT; 81 AA.
AC Q73ZQ7;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein.
GN OrderedLocusNames=MAPI544;
OS Mycobacterium paratuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;
OC Mycobacterium avium complex (MAC).
OX NCBI_TaxID=1770;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=K10;
RA Li L., Bannantine J., Zhang Q., Amonsin A., Alt D., Kapur V.;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR ENBL; AB017232; AA03861.1; -; Genomic_DNA.
KW Complete proteome.
SQ SEQUENCE 81 AA; 8824 MW; D9CCDB7C5A76FCEC CRC64;
```

```
Query Match      100.0%; Score 27; DB 2; Length 81;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GRRGR 5
Db      3 GRRGR 7

RESULT 49
Y696 BORBU      STANDARD;      PRT;      82 AA.
AC OS1639;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Hypothetical UPF0109 protein BB0696.
GN OrderedLocusNames=BB0696;
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=139;
RN [1]_
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685; DOI=10.1038/37551;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R.J., Hickey E.K.,
RA Winn M.L., Dougherty B.A., Tomb J.-F., Fleischmann R.D.,
RA Richardson D.L., Peterson J.D., Karlavage A.R., Quackenbush J.,
RA Salzberg S.L., Hanson M., van Vugt R., Palmer N., Adams M.D.,
RA Gocayne J.D., Weidman J.F., Utterback T.R., Watthey L., McDonald L.A.,
RA Artiach P., Bowman C., Garland S.A., Fujii C., Cotton M.D., Horst K.,
RA Roberts K.M., Hatch B., Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
burgdorferi."
RL Nature 390:580-586(1997).
CC -!- SIMILARITY: Belongs to the UPF0109 family.
CC -!- SIMILARITY: Contains 1 KH domain.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
DR EMBL; AE001170; AAC67047.1; -; Genomic_DNA.
DR PIR; G70186; G70186.
DR TIGR; BB0696; -.
DR HAMAP; MF_00088; -; 1.
DR InterPro; IPR004088; KH_type_1.
DR Pfam; PF00013; KH_1; 1.
DR PROSITE; PS50084; KH_type_1; 1.
KW Complete proteome; Hypothetical protein; RNA-binding.
FT DOMAIN 39 67 KH.
SQ SEQUENCE 82 AA; 9253 MW; 38943030921946B6 CRC64;

Query Match      100.0%; Score 27; DB 1; Length 82;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GRRGR 5
Db      52 GRRGR 56

RESULT 50
Q4VXZ3 HUMAN
ID Q4VXZ3 HUMAN PRELIMINARY;      PRT;      82 AA.
AC Q4VXZ3;
DT 13-SEP-2005 (TReMBLrel. 31, Created)
DT 13-SEP-2005 (TReMBLrel. 31, Last sequence update)
```

```
DT 13-SEP-2005 (TReMBLrel. 31, Last annotation update)
DE Ribosomal protein S18.
GN Name=RPS18; ORFNames=RP5-1033B10.18-002;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Tubby B.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the ribosomal protein S13P family.
DR EMBL; AL031228; CAI95617.1; -; Genomic_DNA.
DR InterPro; IPR001892; Ribosomal_S13.
DR Pfam; PF00416; Ribosomal_S13; 1.
DR ProDom; PD001363; Ribosomal_S13; 1.
DR PROSITE; PS00646; RIBOSOMAL_S13_1; 1.
DR PROSITE; PS50159; RIBOSOMAL_S13_2; 1.
KW Ribonucleoprotein; Ribosomal protein.
SQ SEQUENCE 82 AA; 9767 MW; BD9A4701784DECC1 CRC64;

Query Match      100.0%; Score 27; DB 2; Length 82;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GRRGR 5
Db      70 GRRGR 74

RESULT 51
Q660H5 BORGA
ID Q660H5 BORGA PRELIMINARY;      PRT;      82 AA.
AC Q660H5;
DT 25-OCT-2004 (TReMBLrel. 28, Created)
DT 25-OCT-2004 (TReMBLrel. 28, Last sequence update)
DE Hypothetical protein.
GN OrderedLocusNames=BG0719;
OS Borrelia garinii.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
OC Borrelia burgdorferi group.
OX NCBI_TaxID=29519;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PBI;
RA Gloeckner G., Lehmann R., Romualdi A., Pradella S.,
RA Schulte-Spechtel U., Wilske B., Suehnell J., Platzner M.;
RT "Comparative analysis of the Borrelia garinii genome."
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CP000013; AAU07546.1; -; Genomic_DNA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR004088; KH_type_1.
DR Pfam; PF00013; KH_1; 1.
DR PROSITE; PS50084; KH_TYPE_1; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 82 AA; 9252 MW; 25543030860D46F2 CRC64;

Query Match      100.0%; Score 27; DB 2; Length 82;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GRRGR 5
Db      52 GRRGR 56

RESULT 52
Q6H6S6 ORYSA
ID Q6H6S6 ORYSA PRELIMINARY;      PRT;      84 AA.
AC Q6H6S6;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
```

```

DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Hypothetical protein P0048B08.13 (Hypothetical protein
GN P0519A12.44)
GN Name=P0048B08.13; Synonyms=P0519A12.44;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 2, PAC
RT clone:P0048B08."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 2, PAC
RT clone:P0519A12."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP004868; BAD25573.1; -; Genomic_DNA.
DR EMBL; AP004839; BAD25537.1; -; Genomic_DNA.
DR Gramene; Q6H835; -.
DR InterPro; IPR005829; Sug_transporter.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 84 AA; 8681 MW; 0A1635E815B0F3F8 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 84;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5
DB 55 GRGR 59

RESULT 53
O48451_BSPSP
ID O48451_BSPSP PRELIMINARY; PRT; 85 AA.
AC O48451;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Complete nucleotide sequence.
OS Bacteriophage SPPI.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC Lambda-like viruses.
OX NCBI_TaxID=10724;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Alonso J.C., Luder G., Stiege A.C., Chai S., Weise F., Trautner T.A.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; X97918; CAA66552.1; -; Genomic_DNA.
DR PIR; T42291; T42291.
SQ SEQUENCE 85 AA; 9933 MW; 025E0CFB2E5E3AF CRC64;

Query Match 100.0%; Score 27; DB 2; Length 85;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5
DB 4 GRGR 8

RESULT 54
Q6H835_ORYSA
ID Q6H835_ORYSA PRELIMINARY; PRT; 85 AA.
AC Q6H835;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)

```

```

DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein OJ1267_F10.18.
GN Name=OJ1267_F10.18;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 2, BAC
RT clone:OJ1267_F10."
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP004085; BAD25114.1; -; Genomic_DNA.
DR Gramene; Q6H835; -.
KW Hypothetical protein.
SQ SEQUENCE 85 AA; 8813 MW; 00A85B9055097C3B1 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 85;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5
DB 81 GRGR 85

RESULT 55
Q801B9_LATME
ID Q801B9_LATME PRELIMINARY; PRT; 85 AA.
AC Q801B9;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE HoxB6 (Fragment).
OS Latimeria menadoensis (Indonesian coelacanth).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Coelacanthiformes; Coelacanthidae; Latimeria.
OX NCBI_TaxID=106881;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22457206; PubMed=12547909; DOI=10.1073/pnas.0237317100;
RA Koh E.G., Lam K., Christoffels A., Erdmann M.V., Brenner S.,
RA Venkatesh B.;
RT "Hox gene clusters in the Indonesian coelacanth, Latimeria
RT menadoensis."
RL Proc. Natl. Acad. Sci. U.S.A. 100:1084-1088 (2003).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Koh E.G.L., Lam K., Christoffels A., Erdmann M.V., Brenner S.,
RA Venkatesh B.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
CC -1- AXCELLULAR LOCATION: Nuclear (By similarity).
DR EMBL; AY183738; AAO43031.1; -; Genomic_DNA.
DR HSSP; P02833; 9ANT.
DR SMR; Q801B9; 8-70.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR012287; Homeodomain-rel.
DR InterPro; IPR000047; HTH_lambrepresr.
DR Pfam; PF00046; Homeobox_1.
DR PRINTS; PR00024; HOMEBOX.
DR PRINTS; PR00031; HTHREPRESSR.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOK; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
FT NON_TER 1

```

OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC	Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Labroidel;
OC	Cichlidae; African cichlids; Pseudocrenilabrinae; Tilapiini;
OC	Oreochromis.
OC	NCBI_TaxID=8128;
RN	[1]_ NCBI TaxID=8128;
RP	NUCLEOTIDE SEQUENCE.
RC	STRAIN=B6b;
RA	PubMed=15716008; DOI=10.1016/j.gene.2004.10.027;
RX	Santini S., Bernardi G.;
RT	"Organization and base composition of tilapia Hox genes: implications
RT	for the evolution of Hox clusters in fish.";
RL	Gene 346:51-61(2005).
CC	!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC	EMBL; AY575735; AAY23655.1; -; Genomic_DNA.
DR	SMR; Q533W6; 8-67.
DR	GO; GO:0005634; C:nucleus; IEA.
DR	GO; GO:0003700; E:transcription factor activity; IEA.
DR	GO; GO:0006352; F:regulation of transcription, DNA-dependent; IEA.
DR	InterPro; IPR001356; Homeobox.
DR	Pfam; PF00046; Homeobox; 1.
DR	PRINTS; PR0024; HOMEBOX.
DR	ProDom; PD000010; Homeobox; 1.
DR	SMART; SM00389; HOX; 1.
DR	PROSITE; PS00027; HOMEBOX_1; 1.
DR	PROSITE; PS0071; HOMEBOX_2; 1.
KW	DNA-binding; Homeobox; Nuclear protein.
KW	DNA-binding; Homeobox; Nuclear protein.
FT	NON_TER 1
FT	SEQUENCE 85 AA; 10471 MW; 07CABED1A5A04D05 CRC64;
SEQ	
Query Match 100.0%; Score 27; DB 2; Length 85;	
Best Local Similarity 100.0%; Pred. No. 4.6e+02;	
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps	
QY	1 GRRGR 5
Db	7 GRRGR 11
RESULT 58	
Q50112	MYCLE PRELIMINARY; PRT; 86 AA.
ID	Q50112_MYCLE PRELIMINARY; PRT; 86 AA.
AC	Q50112;
DT	01-NOV-1996 (TREMBLrel. 01, Created)
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT	01-NOV-2003 (TREMBLrel. 23, Last annotation update)
DE	U650t.
OS	Mycobacterium leprae.
OS	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC	Corynebacteriineae; Mycobacteriaceae; Mycobacterium.
OC	NCBI_TaxID=1769;
RN	[1]_ NCBI TaxID=1769;
RP	NUCLEOTIDE SEQUENCE.
RA	Smith D.R.;
RA	Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
RL	[2]
RP	NUCLEOTIDE SEQUENCE.
RP	Robison K.;
RL	Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RL	EMBL; U15184; AAA63058.1; -; Genomic DNA.
SR	SEQUENCE 86 AA; 9895 MW; E2F82BBD618B4F9A CRC64;
SEQ	
Query Match 100.0%; Score 27; DB 2; Length 86;	
Best Local Similarity 100.0%; Pred. No. 4.7e+02;	
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps	
QY	1 GRRGR 5
Db	79 GRRGR 83
RESULT 59	
Q4SQ17	TETNG

RESULT 59  
Q4SQI7\_TETNG



QASQ17\_TETNG PRELIMINARY; PRT; 86 AA.  
 QASQ17;  
 AC 13-SEP-2005 (T-EMBLrel. 31, Created)  
 DT 13-SEP-2005 (T-EMBLrel. 31, Last sequence update)  
 DT 13-SEP-2005 (T-EMBLrel. 31, Last annotation update)  
 DE Chromosome 17 SCAF14532, whole genome shotgun sequence.  
 DE (Fragment).  
 GN ORFNames=GSTENG00014355001;  
 OS Tetraodon nigroviridis (Green puffer).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 OC Tetraodontoidea; Tetraodontidae; Tetraodon.  
 OX NCBI\_TaxID=99883;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,  
 RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,  
 RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,  
 RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,  
 RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,  
 RA Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,  
 RA Cruaud C., Duprat S., Brottier P., Couranceau J.P., Gouzy J., Bosak S.,  
 RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,  
 RA Keilts M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,  
 RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,  
 RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,  
 RA Wincker P., Lander E.S., Weissenbach J., Roest Crolius H.;  
 RT "genome duplication in the teleost fish Tetraodon nigroviridis reveals  
 RT the early vertebrate proto-karyotype";  
 RL Nature 431:946-957 (2004).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RG Genoscope; Whitehead Institute Centre for Genome Research;  
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.  
 CC -1- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 DR EMBL; CAEE01014532; CAF97095.1; -; Genomic\_DNA.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 86 AA; 9942 MW; ADFC4D7BA9949687 CRC64;  
 Query Match 100.0%; Score 27; DB 2; Length 86;  
 Best Local Similarity 100.0%; Pred. No. 4.7e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GRGR 5  
 DB 55 GRGR 59  
 RESULT 60  
 Q9LIX6\_ORYSA  
 ID Q9LIX6\_ORYSA PRELIMINARY; PRT; 88 AA.  
 AC Q9LIX6;  
 DT 01-OCT-2000 (T-EMBLrel. 15, Created)  
 DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)  
 DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)  
 DE Hypothetical protein.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Eriarthroideae; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Haing Y.C., Chow T., Chen C., Wu H., Chu M., Chao Y., Liu S.;  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP001111; BAA90493.1; -; Genomic\_DNA.  
 DR Gramene; Q9LIX6; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 88 AA; 9782 MW; FEF89DB1DB53609A CRC64;

Query Match 100.0%; Score 27; DB 2; Length 88;  
 Best Local Similarity 100.0%; Pred. No. 4.8e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GRGR 5  
 DB 50 GRGR 54  
 RESULT 61  
 Q8FRF8\_COREF  
 ID Q8FRF8\_COREF PRELIMINARY; PRT; 89 AA.  
 AC Q8FRF8;  
 DT 01-MAR-2003 (T-EMBLrel. 23, Created)  
 DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 GN OrderedLocusNames=CE0803;  
 OS Corynebacterium efficiens.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
 OX NCBI\_TaxID=152794;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;  
 RX MEDLINE=22723752; PubMed=12840036; DOI=10.1101/gr.1285603;  
 RA Nishio Y., Nakamura Y., Kawarabayashi Y., Usuda Y., Kimura E.,  
 RA Sugimoto S., Matsui K., Yamagishi A., Kikuchi H., Ikeo K.,  
 RA Gojobori T.;  
 RT "Comparative complete genome sequence analysis of the amino acid  
 RT replacements responsible for the thermostability of Corynebacterium  
 RT efficiens";  
 RL Genome Res. 13:1572-1579 (2003).  
 DR EMBL; BA000035; BAC17613.1; -; Genomic DNA.  
 KW Complete proteome; Hypothetical protein.  
 SQ SEQUENCE 89 AA; 9313 MW; 1A3031A3C5D3856B CRC64;  
 Query Match 100.0%; Score 27; DB 2; Length 89;  
 Best Local Similarity 100.0%; Pred. No. 4.9e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GRGR 5  
 DB 76 GRGR 80  
 RESULT 62  
 Q53PI9\_ORYSA  
 ID Q53PI9\_ORYSA PRELIMINARY; PRT; 91 AA.  
 AC Q53PI9;  
 DT 13-SEP-2005 (T-EMBLrel. 31, Created)  
 DT 13-SEP-2005 (T-EMBLrel. 31, Last sequence update)  
 DT 13-SEP-2005 (T-EMBLrel. 31, Last annotation update)  
 DE Hypothetical protein.  
 GN ORFNames=LOC\_Os11g24210;  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartioideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=39947;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Buell R.;  
 RA Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.

DR	EMBL; AC120536; AAX96498.1; -; Genomic_DNA.
KW	Hypothetical protein.
SQ	SEQUENCE 91 AA; 10091 MW; 8F875D1EBF7B5BEA CRC64;
Query Match	100.0%; Score 27; DB 2; Length 91;
Best Local Similarity	100.0%; Pred. No. 5e+02;
Matches	5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 GRGR 5 
Db	36 GRGR 40
RESULT 63	
Q51CD7_ENTHI	PRELIMINARY; PRT; 92 AA.
ID	Q51CD7;
AC	Q51CD7;
DT	13-SEP-2005 (TrEMBLrel. 31, Created)
DT	13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DE	40S ribosomal protein S14, putative.
GN	ORFNAMES=21.t00010;
OS	Entamoeba histolytica HM-1:IMSS.
OC	Eukaryota; Entamoebidae; Entamoeba.
NCBI_TaxID=294381;	
[1]	
NUCLEOTIDE SEQUENCE.	
STRAIN=HM-1;IMSS;	
PUBMED=15729342; DOI=10.1038/nature03291;	
Loftus B., Anderson I., Davies R., Alsmark U.C., Samuelson J., Whitehead S., Amedeo P., Roncaglia P., Bertram M., Hirt R.P., Mann B.J., Nozaki T., Suh B., Pop M., Duchene M., Ackers J., Tannich E., Leippe M., Hofer M., Bruchhaus I., Willhoest U., Bhattacharya A., Chillingworth T., Churcher C., Hance Z., Harris B., Harris D., Jagels K., Moule S., Mungall K., Ormond D., Squares R., Whitehead S., Quail M.A., Rabinowitsch E., Norbertczak H., Price C., Wang Z., Guillen N., Gilchrist C., Stroup S.E., Bhattacharya S., Lohia A., Foster P.G., Sicheritz-Ponten T., Weber C., Singh U., Mukherjee C., El-Sayed N.M., Petri W.A., Clark C.G., Embley T.M., Barrell B., Fraser C.M., Hall N.;	
"The genome of the protist parasite Entamoeba histolytica."	
Nature 433:865-868 (2005).	
-!- CAUTION: The sequence shown here is derived from an EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is preliminary data.	
-!- SIMILARITY: Belongs to the ribosomal protein S11P family.	
EMBL; AAF01000095; EAL50513.1; -; Genomic DNA.	
GO: GO:0003735; F:structural constituent of ribosome; IEA.	
InterPro: IPR001971; Ribosomal_S11.	
Pfam: PF00411; Ribosomal_S11; 1.	
ProDom: PD001010; Ribosomal_S11; 1.	
PROSITE: PS00054; RIBOSOMAL_S11; UNKNOWN_1.	
Ribonucleoprotein; Ribosomal protein.	
SQ	SEQUENCE 92 AA; 9939 MW; 5CB472D0265512BD CRC64;
Query Match	100.0%; Score 27; DB 2; Length 92;
Best Local Similarity	100.0%; Pred. No. 5e+02;
Matches	5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 GRGR 5 
Db	86 GRGR 90
RESULT 64	
Q5CMB2_CRYHO	PRELIMINARY; PRT; 92 AA.
ID	Q5CMB2;
AC	Q5CMB2;
DT	10-MAY-2005 (TrEMBLrel. 30, Created)
DT	10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DE	40S ribosomal protein S14.
GN	ORFNAMES=Chro.70025;
OS	Cryptosporidium hominis.
OC	Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
OC	Cryptosporidiidae; Cryptosporidium.
NCBI_TaxID=237895;	
[1]	
NUCLEOTIDE SEQUENCE.	
STRAIN=TU502;	
Xu P., Widmer G., Wang Y., Ozaki L.S., Alves J.M., Serrano M.G., Dear P.H., Fuli D., Manque P., Akiyoshi D., Mackey A.J., Pearson W.R., Dear P.H., Bankier A.T., Peterson D.L., Abrahamsen M.S., Kapur V., Tzipori S., Buck G.A.;	
"The genome of Cryptosporidium hominis."	
Nature 431:1107-1112(2004).	
-!- SIMILARITY: Belongs to the ribosomal protein S11P family.	
EMBL; AAELO1000031; EAL37752.1; -; Genomic DNA.	
SMR; Q5CMB2; 1-88.	
GO: GO:0005840; C:ribosome; IEA.	
GO: GO:0003735; F:structural constituent of ribosome; IEA.	
GO: GO:0006412; P:protein biosynthesis; IEA.	
InterPro: IPR001971; Ribosomal_S11.	
Pfam: PF00411; Ribosomal_S11; 1.	
PROSITE: PS00054; RIBOSOMAL_S11; 1.	
Ribonucleoprotein; Ribosomal protein.	
SQ	SEQUENCE 92 AA; 9929 MW; 06C98BBDF3CDDFFD CRC64;
Query Match	100.0%; Score 27; DB 2; Length 92;
Best Local Similarity	100.0%; Pred. No. 5e+02;
Matches	5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 GRGR 5 
Db	86 GRGR 90
RESULT 65	
Q8WVG6_HUMAN	PRELIMINARY; PRT; 93 AA.
ID	Q8WVG6;
AC	Q8WVG6;
DT	01-WAR-2002 (TrEMBLrel. 20, Created)
DT	01-WAR-2002 (TrEMBLrel. 20, Last sequence update)
DT	01-WAR-2004 (TrEMBLrel. 26, Last annotation update)
DE	Hypothetical protein (Fragment).
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae;
OC	Homo.
NCBI_TaxID=9606;	
[1]	
NUCLEOTIDE SEQUENCE.	
Tissue=Brain;	
Strasbourg R.;	
Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.	
EMBL; BC018035; AAH18035.1; -; mRNA.	
Ensembl; ENSG0000178743; Homo sapiens.	
Hypothetical protein.	
NON TER 1	
SQ	SEQUENCE 93 AA; 10739 MW; 794978AEAE0B04F CRC64;
Query Match	100.0%; Score 27; DB 2; Length 93;
Best Local Similarity	100.0%; Pred. No. 5.1e+02;
Matches	5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 GRGR 5 
Db	18 GRGR 22
RESULT 66	
Q5N9T5_ORYSA	PRELIMINARY; PRT; 93 AA.
ID	Q5N9T5;
AC	Q5N9T5;
DT	01-FEB-2005 (TrEMBLrel. 29, Created)
DT	01-FEB-2005 (TrEMBLrel. 29, Last sequence update)

```

DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Hypothetical protein P0414E03.10.
GN Names=P0414E03.10;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K., Sakata K., Baba T., Katayose Y.,
RA Wu J., Niimura Y., Cheng Z., Nagamura Y., Antonio B.A., Kanamori H.,
RA Hosokawa S., Masukawa M., Arikawa K., Chiden Y., Hayashi M.,
RA Okamoto M., Ando T., Aoki H., Arita K., Hamada M., Harada C.,
RA Hijishita S., Honda M., Ichikawa Y., Idonuma A., Iijima M., Ikeda M.,
RA Ikono M., Itoh S., Itoh Y., Itoh Y., Iwabuchi A., Kamiya K.,
RA Karasawa W., Katagiri S., Kikuta A., Kobayashi N., Kono I.,
RA Machita K., Maehara T., Mizuno H., Mizubayashi T., Mukai Y.,
RA Nagasaki H., Nakashima M., Nakama Y., Nakamichi Y., Nakamura M.,
RA Namiki N., Negishi M., Ohta I., Ono N., Saji S., Sakai K., Shibata M.,
RA Shimokawa T., Shomura A., Song J., Takazaki Y., Terasawa K., Tsuji K.,
RA Waki K., Yamagata H., Yamane H., Yoshiki S., Yoshihara R., Yukawa K.,
RA Zhong H., Iwama H., Endo T., Ito H., Hahn J.H., Kim H.I., Eun M.Y.,
RA Yano M., Jiang J., Gojobori T.;
RT "The genome sequence and structure of rice chromosome 1."
RL Nature 420:312-316(2002).
DR EMBL; AP003242; BAD81772.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 93 AA; 10182 MW; D091DB95D953AD20 CRC64;
Query Match 100.0%; Score 27; DB 2; Length 93;
Best Local Similarity 100.0%; Pred. No. 5.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GRRGR 5
DB 28 GRRGR 32
RESULT 67
Q43J30 AZOVI
ID Q43J30 AZOVI PRELIMINARY; PRT; 93 AA.
AC Q43J30;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORFNames=AvindRAFT_7757;
OS Azotobacter vinelandii AvOP.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Azotobacter.
OX NCBI_TaxID=322710;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AVOP;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome assembly of Azotobacter vinelandii
RT AvOP."
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AVOP;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Azotobacter vinelandii
RT AvOP."
RN Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AVOP;
RG US DOE Joint Genome Institute;
RA

```

```

RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AVOP;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AAU03000003; EAM07851.1; -; Genomic_DNA.
DR Hypothetical protein.
SQ SEQUENCE 93 AA; 10591 MW; 29EC09C240CB497B CRC64;
Query Match 100.0%; Score 27; DB 2; Length 93;
Best Local Similarity 100.0%; Pred. No. 5.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GRRGR 5
DB 10 GRRGR 14
RESULT 68
Q86MX2 DICDI
ID Q86MX2 DICDI PRELIMINARY; PRT; 94 AA.
AC Q86MX2;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE CoMh (Fragment).
GN Names=comH;
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Kibler K., Nguyen T.-L., Svetz J., Van Driessche N., Ibarra M.,
RA Thompson C., Shaw C., Shaulsky G.;
RT "A novel developmental mechanism in Dictyostelium revealed in a screen
RT for communication mutants."
RL Dev. Biol. 0:0-0(2003).
DR EMBL; AY221645; AA034401.1; -; Genomic_DNA.
DR DictyBase; DDB0214839; comH.
FT NON TER 94 94
SQ SEQUENCE 94 AA; 9519 MW; 70316F9F75E8C257 CRC64;
Query Match 100.0%; Score 27; DB 2; Length 94;
Best Local Similarity 100.0%; Pred. No. 5.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GRRGR 5
DB 74 GRRGR 78
RESULT 69
Q5NB13 ORYSA
ID Q5NB13 ORYSA PRELIMINARY; PRT; 94 AA.
AC Q5NB13;
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Hypothetical protein P0511C01.14.
GN Name=P0511C01.14;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.

```

RA Saeaki T., Matsumoto T., Yamamoto K., Sakata K., Baba T., Katayose Y.,  
RA Wu J., Nilmura Y., Cheng Z., Nagamura Y., Antonio B.A., Kanamori H.,  
RA Hosokawa S., Masukawa M., Arikawa K., Chiden Y., Hayashi M.,  
RA Okamoto M., Ando T., Aoki H., Arita K., Hamada M., Harada C.,  
RA Hijishita S., Honda M., Ichikawa Y., Idonuma A., Iijima M., Ikeda M.,  
RA Ikono M., Itoh T., Itoh Y., Iwabu A., Iwabuchi A., Kamiya K.,  
RA Karasawa W., Katagiri S., Kikuta A., Kobayashi N., Kono I.,  
RA Machata K., Maehara T., Mizuno H., Mizubayashi T., Mukai Y.,  
RA Nagasaki H., Nakashima M., Nakama Y., Nakamichi Y., Nakamura M.,  
RA Namiki N., Negishi M., Ohta I., Ono N., Saji S., Sakai K., Shibata M.,  
RA Shinokawa T., Shomura A., Song J., Takazaki Y., Terasawa K., Tsuji K.,  
RA Waki K., Yamagata H., Yamane H., Yoshiki S., Yoshihara R., Yukawa K.,  
RA Zhong H., Iwana H., Endo T., Ito H., Hahn J.H., Kim H.I., Eun M.Y.,  
RA Yano M., Jiang J., Gojobori T.,  
RT "The genome sequence and structure of rice chromosome 1.";  
RL Nature 420:312-316 (2002).  
DR EMBL; AF002070; BAD81339.1; -; Genomic\_DNA.  
KW Hypothetical protein.  
SQ SEQUENCE 94 AA; 10435 MW; 2DD94556D67C7F7F CRC64;  
  
Query Match 100.0%; Score 27; DB 2; Length 94;  
Best Local Similarity 100.0%; Pred. No. 5.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GRRGR 5  
DB 24 GRRGR 28  
  
RESULT 70  
Q7SUV3\_9HIV1 PRELIMINARY; PRT; 94 AA.  
AC Q7SUY3;  
DT 01-OCT-2003 (TREMBlrel. 25, Created)  
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)  
DE 01-WAR-2004 (TREMBlrel. 26, Last annotation update)  
DE Vpr protein.  
GN Name=vpr;  
OS Human immunodeficiency virus 1.  
OC Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;  
OC Lentivirus; Primate lentivirus group.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RA Takebe Y., Motomura K., Tatsumi M., Lwin H., Zaw M., Kusagawa S.;  
RA Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB097872; BAC77505.1; -; Genomic\_DNA.  
DR HSSP; P12520; 1DSJ.  
DR SMR; Q7SUY3: 1-94.  
DR InterPro; IPR000012; Retrov\_Vpr/X.  
DR Pfam; PF00522; VPR; 1.  
DR PRINTS; PR00444; HIVVPRVX.  
KW AIDS.  
SQ SEQUENCE 94 AA; 10928 MW; 5D79BD6A246414B3 CRC64;  
  
Query Match 100.0%; Score 27; DB 2; Length 94;  
Best Local Similarity 100.0%; Pred. No. 5.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GRRGR 5  
DB 84 GRRGR 88  
  
RESULT 71  
Q8UMG3\_9HIV1 PRELIMINARY; PRT; 94 AA.  
AC Q8UMG3;  
DT 01-MAR-2002 (TREMBlrel. 20, Created)  
DT 01-WAR-2002 (TREMBlrel. 20, Last sequence update)  
DT 01-WAR-2004 (TREMBlrel. 26, Last annotation update)  
DE Vpr protein.  
GN Name=Vpr;

OS Human immunodeficiency virus 1.  
OC Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;  
OC Lentivirus; Primate lentivirus group.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=21828193; PubMed=11839159; DOI=10.1089/08992220252781301;  
RA Montavon C., Vergne L., Bourgeois A., Mpoudi-Erogle E.,  
RA Malonga-Mouellet G., Butel C., Toure-Kane C., Delaporte E.,  
RA Peeters M.;  
RT "Identification of a new circulating recombinant form of HIV type 1,  
RT CRF11-cpx, involving subtypes A, G, J, and CRF01-AG, in Central  
RT Africa.";  
RL AIDS Res. Hum. Retroviruses 18:231-236 (2002).  
DR EMBL; AJ291720; CAC88004.1; -; Genomic\_DNA.  
DR HSSP; P12520; 1DSJ.  
DR SMR; Q8UMG3: 1-93.  
DR InterPro; IPR000012; Retrov\_Vpr/X.  
DR Pfam; PF00522; VPR; 1.  
DR PRINTS; PR00444; HIVVPRVX.  
KW AIDS.  
SQ SEQUENCE 94 AA; 10963 MW; F790DB16B7BD1403 CRC64;  
  
Query Match 100.0%; Score 27; DB 2; Length 94;  
Best Local Similarity 100.0%; Pred. No. 5.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GRRGR 5  
DB 84 GRRGR 88  
  
RESULT 72  
Q6EQW9\_ORYSA PRELIMINARY; PRT; 95 AA.  
AC Q6EQW9;  
DT 25-OCT-2004 (TREMBlrel. 28, Created)  
DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)  
DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)  
DE Hypothetical protein OSUNBA0086N11.19;  
GN Name=OSUNBA0086N11.19;  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Magnoliophyta; Liliopsida; Poaceae;  
OC Eukaryota; Viridiplantae; Magnoliophyta; Liliopsida; Poaceae;  
OC Eukaryota; Viridiplantae; Magnoliophyta; Liliopsida; Poaceae;  
OX NCBI\_TaxID=39947;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RA Sasaki T., Matsumoto T., Katayose Y.;  
RA "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 2, BAC  
RT clone:OSJNBA0086N11.";  
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF005651; BAD28951.1; -; Genomic\_DNA.  
KW Hypothetical protein.  
SQ SEQUENCE 95 AA; 9338 MW; B4ACB8E694D4FD16 CRC64;  
  
Query Match 100.0%; Score 27; DB 2; Length 95;  
Best Local Similarity 100.0%; Pred. No. 5.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GRRGR 5  
DB 59 GRRGR 63  
  
RESULT 73  
Q9Z291\_9RODE PRELIMINARY; PRT; 95 AA.  
AC Q9Z291;  
DT 01-MAY-1999 (TREMBlrel. 10, Created)  
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

```
DE Ribosomal protein S18 (Fragment).
OS Cricetus sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Cricetinae; Cricetulus.
OX NCBI_TaxID=10031;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Bachrati C.Z., Downes S., Raako I.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the ribosomal protein S13p family.
DR ENBL; AF081143; AAD03879.1; -; mRNA.
DR GO; GO:0005840; C:ribosome; IEA.
DR GO; GO:0003723; F:rRNA binding; IEA.
DR GO; GO:0003735; F:protein biosynthesis; IEA.
DR GO; GO:0006412; P:protein biosynthesis; IEA.
DR InterPro; IPR001892; Ribosomal_S13.
DR Pfam; PF00416; Ribosomal_S13; 1.
DR ProDom; PD001363; Ribosomal_S13; 1.
DR PROSITE; PS00646; RIBOSOMAL_S13_1; 1.
DR PROSITE; PS0159; RIBOSOMAL_S13_2; 1.
KW Ribonucleoprotein; Ribosomal protein.
FT NON TER 1
SQ SEQUENCE 95 AA; 11307 MW; 22B8959B4F28D04C CRC64;

Query Match 100.0%; Score 27; DB 2; Length 95;
Best Local Similarity 100.0%; Pred. No. 5.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5
DB 83 GRGR 87

RESULT 74
Q6WS73_9HIV1
ID Q6WS73_9HIV1 PRELIMINARY; PRT; 95 AA.
AC Q6WS73;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Vpr protein.
GN Names=vpr;
OS Human immunodeficiency virus 1.
OC Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;
OC Lentivirus; Primate lentivirus group.
OX NCBI_TaxID=11676;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA MEDLINE=22836537; PubMed=12954230; DOI=10.1016/S0042-6822(03)00381-7;
RA Tebit D.M., Zekeng L., Kaptue L., Krausslich H.-G., Herchenroder O.;
RT "Construction and characterization of a full-length infectious
RT molecular clone from a fast replicating, X4-tropic HIV-1 CRF02_AG
RT primary isolate.";
RL Virology 313:645-652(2003).
DR ENBL; AY271690; AAQ17101.1; -; Genomic_DNA.
DR HSSP; P12520; 1BDE.
DR SMR; Q6WS73; 1-95.
DR InterPro; IPR000012; Retrov_Vpr/X.
DR Pfam; PF00522; VPR; 1.
DR PRINTS; PR00444; HIVVPRVPX.
KW AIDS.
SQ SEQUENCE 95 AA; 11043 MW; 324893D951A836B3 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 95;
Best Local Similarity 100.0%; Pred. No. 5.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5
DB 85 GRGR 89
```

```
RESULT 75
Q97061_9HIV1
ID Q97061_9HIV1 PRELIMINARY; PRT; 95 AA.
AC Q97061;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Vpr.
GN Names=vpr;
OS Human immunodeficiency virus 1.
OC Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;
OC Lentivirus; Primate lentivirus group.
OX NCBI_TaxID=11676;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA MEDLINE=96323109; PubMed=8709215;
RA Carr J.K., Salminen M.O., Koch C., Gotte D., Arstenstein A.W.,
RA Hegerich P.A., St Louis D., Burke D.S., McCutchan F.E.;
RT "Full-length sequence and mosaic structure of a human immunodeficiency
RT virus type 1 isolate from Thailand.";
RL J. Virol. 70:5935-5943(1996).
DR ENBL; U54771; AAC55462.1; -; Genomic_DNA.
DR HSSP; P05954; 1FI0.
DR SMR; Q97061; 1-95.
DR InterPro; IPR000012; Retrov_Vpr/X.
DR Pfam; PF00522; VPR; 1.
DR PRINTS; PR00444; HIVVPRVPX.
KW AIDS.
SQ SEQUENCE 95 AA; 11168 MW; 481EF2F928798CEE CRC64;

Query Match 100.0%; Score 27; DB 2; Length 95;
Best Local Similarity 100.0%; Pred. No. 5.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRGR 5
DB 85 GRGR 89

Search completed: December 2, 2005, 10:07:13
Job time : 183 secs
```

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2005, 09:42:09 / Search time 13 Seconds  
(without alignments)  
37.006 Million cell updates/sec

Title: SEQ-RGRGR  
Perfect score: 27  
Sequence: 1 rgrgr 5

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 1000 summaries

Database : PIR\_80:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	27	100.0	55	2	T36132
2	27	100.0	59	2	I54837
3	27	100.0	69	2	S39424
4	27	100.0	90	2	JC5931
5	27	100.0	95	2	H72271
6	27	100.0	96	2	JC5933
7	27	100.0	96	2	A31895
8	27	100.0	99	2	D97652
9	27	100.0	100	2	T44039
10	27	100.0	105	2	JC5932
11	27	100.0	107	2	A32794
12	27	100.0	107	2	T49594
13	27	100.0	108	2	JC4575
14	27	100.0	108	2	AC0808
15	27	100.0	109	2	AC2232
16	27	100.0	110	1	F71129
17	27	100.0	110	2	F75034
18	27	100.0	110	2	T17833
19	27	100.0	112	2	S57448
20	27	100.0	115	2	D69188
21	27	100.0	116	2	B85036
22	27	100.0	117	2	T38440
23	27	100.0	119	2	A27668
24	27	100.0	121	2	T40190
25	27	100.0	124	2	D71355
26	27	100.0	125	2	A32484
27	27	100.0	126	2	T16952
28	27	100.0	129	2	F97593
29	27	100.0	129	2	AC2815

30	27	100.0	129	2	T46388
31	27	100.0	136	2	T31498
32	27	100.0	137	2	T43028
33	27	100.0	139	2	A38612
34	27	100.0	141	2	T39554
35	27	100.0	142	1	BVECTR
36	27	100.0	142	2	AF0592
37	27	100.0	142	2	F85576
38	27	100.0	142	2	E90725
39	27	100.0	152	2	S22311
40	27	100.0	154	2	T04538
41	27	100.0	159	2	G72249
42	27	100.0	166	2	T39586
43	27	100.0	166	2	T18691
44	27	100.0	170	2	S43476
45	27	100.0	171	2	AC2855
46	27	100.0	176	2	S22310
47	27	100.0	177	2	A71011
48	27	100.0	177	2	T49691
49	27	100.0	178	2	T09585
50	27	100.0	178	2	T09584
51	27	100.0	183	2	T39529
52	27	100.0	184	1	S47020
53	27	100.0	184	2	H97631
54	27	100.0	186	2	C87500
55	27	100.0	189	2	A48834
56	27	100.0	194	2	C72450
57	27	100.0	194	2	D83175
58	27	100.0	197	2	S57948
59	27	100.0	198	2	T19797
60	27	100.0	199	2	T48099
61	27	100.0	201	2	G81967
62	27	100.0	201	2	B81024
63	27	100.0	202	1	D69321
64	27	100.0	203	2	T24537
65	27	100.0	205	2	B83495
66	27	100.0	206	2	JC1235
67	27	100.0	208	2	T21689
68	27	100.0	210	2	C84404
69	27	100.0	210	2	B75450
70	27	100.0	210	2	T41982
71	27	100.0	210	2	A32398
72	27	100.0	211	2	C96539
73	27	100.0	213	2	T03931
74	27	100.0	215	2	T49743
75	27	100.0	223	2	T19793
76	27	100.0	243	2	A96744
77	27	100.0	249	2	T16446
78	27	100.0	252	2	AC0869
79	27	100.0	253	2	T41418
80	27	100.0	256	2	AD3235
81	27	100.0	261	2	H69122
82	27	100.0	261	2	AB3070
83	27	100.0	264	2	H70434
84	27	100.0	265	2	B83395
85	27	100.0	267	2	S30395
86	27	100.0	268	2	D71657
87	27	100.0	269	2	B47112
88	27	100.0	272	2	B72641
89	27	100.0	272	2	S73796
90	27	100.0	272	2	S59042
91	27	100.0	276	2	T46185
92	27	100.0	281	2	S65692
93	27	100.0	282	2	T52466
94	27	100.0	282	2	T51013
95	27	100.0	282	2	F96687
96	27	100.0	284	2	T50673
97	27	100.0	286	2	T31494
98	27	100.0	286	2	T09704
99	27	100.0	289	2	C84173
100	27	100.0	290	2	T45402
101	27	100.0	293	1	R3RTS2
102	27	100.0	294	2	T19920

hypothetical prote  
hypothetical prote  
HMG-protein I alph  
insulin-like growt  
very hypothetical  
tolR protein - Esc  
tolR protein (impo  
probable inner mem  
probable inner mem  
HMG-Y-related prot  
hypothetical prote  
ribosomal protein  
rna binding protei  
hypothetical prote  
histone-like DNA-b  
ECF family sigma f  
HMG-Y-related prot  
hypothetical prote  
hypothetical prote  
high mobility grou  
high mobility grou  
hypothetical prote  
ribosomal protein  
hypothetical prote  
conserved hypoteth  
basic fibroblast g  
probable DNA-3-met  
hypothetical prote  
HMG/Y protein - g  
hypothetical prote  
hypothetical prote  
probable regulator  
stringent starvati  
conserved hypoteth  
hypothetical prote  
conserved hypoteth  
transcription fact  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote  
40s ribosomal prot  
2-deoxy-D-gluconat  
biotin acetyl-CoA  
conserved hypoteth  
conserved hypoteth  
probable enoyl-CoA  
ribosomal protein  
biotin-protein lig  
growth response pr  
hypothetical prote  
hypothetical prote  
splicing factor SR  
ribosomal protein  
B4 protein - human  
hypothetical prote  
related to calmodu  
hypothetical prote  
ribosomal protein  
hypothetical prote  
probable arginine/  
heat shock protein  
hypothetical prote  
ribosomal protein  
hypothetical prote

103	27	100.0	295	4	S36174	RNA binding protei	176	27	100.0	483	2	S36470	E2 protein - human
104	27	100.0	296	2	S52254	copper resistance	177	27	100.0	487	2	S42442	nuclear protein EB
105	27	100.0	297	2	T01458	hypothetical prote	178	27	100.0	488	2	S36570	E2 prote in - human
106	27	100.0	300	2	T43225	hypothetical prote	179	27	100.0	491	2	H84477	probable PttA-like
107	27	100.0	301	2	JW0079	heterogeneous nucl	180	27	100.0	492	2	JC5169	alkaline nuclease
108	27	100.0	306	2	T19997	hypothetical prote	181	27	100.0	492	2	A41907	methyl-CpG-binding
109	27	100.0	307	2	T34143	hypothetical prote	182	27	100.0	493	2	S36488	E2 protein - human
110	27	100.0	308	2	A29379	glycine-rich prote	183	27	100.0	494	1	A29079	lymphocyte surface
111	27	100.0	308	2	G86464	P12G12.10 protein	184	27	100.0	494	2	A88474	protein C05D10.1 {
112	27	100.0	309	2	G87498	hypothetical prote	185	27	100.0	498	1	W2WL8	E2 protein - human
113	27	100.0	311	2	B84428	hypothetical prote	186	27	100.0	500	2	S55785	nucleolar protein
114	27	100.0	313	2	F72366	methionyl-tRNA for	187	27	100.0	501	2	S49422	11S globulin seed
115	27	100.0	315	2	T43029	HMG protein I beta	188	27	100.0	506	1	W2WL47	E2 protein - human
116	27	100.0	317	2	T00986	yeast pheromone re	189	27	100.0	509	2	S18872	legumin-like stora
117	27	100.0	318	2	F81805	probable invertase	190	27	100.0	510	2	S62901	legumin 3l precurs
118	27	100.0	318	2	E81070	pinin gene inverti	191	27	100.0	511	1	A48560	UL54 protein - hum
119	27	100.0	319	2	H98216	hypothetical prote	192	27	100.0	512	1	WMBEX4	UL54 protein - hum
120	27	100.0	320	2	T09555	fibrillar in - Arab	193	27	100.0	512	1	WMBEX4	UL54 protein - hum
121	27	100.0	321	2	A38712	fibrillar in [valid	194	27	100.0	512	2	E83060	iron (III)-transpo
122	27	100.0	331	2	T01824	hypothetical prote	195	27	100.0	516	1	FWCNBB	beta-globulin B pr
123	27	100.0	335	2	S43339	hypothetical prote	196	27	100.0	517	2	B87478	hypothetical prote
124	27	100.0	335	2	T31600	glycer aldehyde-3-p	197	27	100.0	521	2	E86332	hypothetical prote
125	27	100.0	335	2	B86429	hypothetical prote	198	27	100.0	523	2	E96576	unknown protein, 4
126	27	100.0	337	2	D87122	F26G16.5 protein -	199	27	100.0	528	2	G02127	fus-like protein -
127	27	100.0	341	2	H83416	probable hypothe	200	27	100.0	530	2	D83469	probable helicase
128	27	100.0	341	2	T01812	hypothetical prote	201	27	100.0	531	2	T08760	hypothetical prote
129	27	100.0	344	2	F75421	hypothetical prote	202	27	100.0	533	2	T00742	ubiquitin-binding
130	27	100.0	347	2	T06584	probable DNA-bind	203	27	100.0	533	2	E86412	hypothetical prote
131	27	100.0	348	2	T02462	probable AR-hook D	204	27	100.0	534	2	JC4572	signal recognition
132	27	100.0	348	2	T47923	probable DNA-bind	205	27	100.0	536	2	A40676	sphere organelle p
133	27	100.0	355	2	S41285	coat protein - swe	206	27	100.0	538	2	S57459	hook-containing pr
134	27	100.0	358	2	H84496	probable retroelem	207	27	100.0	538	2	H86335	T20H2.2 protein -
135	27	100.0	364	2	S26458	hypothetical prote	208	27	100.0	551	2	S77447	HPBRII-7 protein -
136	27	100.0	364	2	S26448	hypothetical prote	209	27	100.0	556	2	T49501	hypothetical prote
137	27	100.0	369	2	F96788	protein T4012.22 {	210	27	100.0	557	2	A96527	probable nuM1 prot
138	27	100.0	370	2	T12180	probable transcrip	211	27	100.0	558	2	T40007	Cct7p - fission ye
139	27	100.0	373	2	S52766	secF protein - Str	212	27	100.0	570	2	T00579	probable laccase {
140	27	100.0	374	2	T33328	hypothetical prote	213	27	100.0	576	2	S50113	collin, p80 - huma
141	27	100.0	378	2	D88042	protein F56D12.5 {	214	27	100.0	578	2	G75622	probable oligopept
142	27	100.0	378	2	S12190	mobl protein - Thi	215	27	100.0	578	2	T48795	origin recognition
143	27	100.0	380	1	F69083	GTP-binding protei	216	27	100.0	589	2	T50385	hypothetical colle
144	27	100.0	380	2	T19932	hypothetical prote	217	27	100.0	594	2	G96525	protein TIN15.25 {
145	27	100.0	391	1	TVBE11	44K protein kinase	218	27	100.0	596	1	WNBEH2	UL32 protein - hum
146	27	100.0	397	2	T40539	zinc finger protei	219	27	100.0	596	2	A45195	adenylcylcyclase ty
147	27	100.0	398	2	T46312	hypothetical prote	220	27	100.0	607	2	S60658	legumin - Gnetum g
148	27	100.0	404	2	T05553	hypothetical prote	221	27	100.0	611	2	T04510	hypothetical prote
149	27	100.0	406	2	A54857	transcription fact	222	27	100.0	611	2	B86387	unknown protein {i
150	27	100.0	407	2	F85079	probable transposo	223	27	100.0	616	2	T42571	gene 28 protein -
151	27	100.0	415	2	T32490	hypothetical prote	224	27	100.0	621	2	S55330	fragile X mental r
152	27	100.0	423	2	A70947	hypothetical prote	225	27	100.0	632	2	T00084	hypothetical prote
153	27	100.0	430	2	A56209	transcription fact	226	27	100.0	633	2	T02673	heterogeneous nucl
154	27	100.0	437	2	JC4929	transcription fact	227	27	100.0	634	2	T17232	hypothetical prote
155	27	100.0	438	1	ERADC2	early E1B 49K prot	228	27	100.0	637	2	S35221	globulin Bgl prec
156	27	100.0	440	2	G87354	hypothetical prote	229	27	100.0	641	1	Q0BE31	nuclear antigen BB
157	27	100.0	443	2	P00450	tryptophan synthas	230	27	100.0	644	2	A53184	myc far upstream e
158	27	100.0	447	2	S53982	hypothetical prote	231	27	100.0	648	2	A57284	spermatid perinucl
159	27	100.0	448	1	S24756	vicilin-like stora	232	27	100.0	653	2	C70742	probable rsbU prot
160	27	100.0	451	2	H75593	hypothetical prote	233	27	100.0	654	2	S76870	hypothetical prote
161	27	100.0	452	2	T39761	probable respirato	234	27	100.0	655	1	A55726	RNA-binding protei
162	27	100.0	456	2	D84651	hypothetical prote	235	27	100.0	656	1	A49358	RNA-binding protei
163	27	100.0	459	2	JC5139	vitronectin precur	236	27	100.0	656	2	D96831	hypothetical prote
164	27	100.0	460	2	T19062	hypothetical prote	237	27	100.0	670	2	S22293	zinc finger protei
165	27	100.0	465	1	W2BEB4	51.3K capsid prote	238	27	100.0	673	2	S60173	fragile X mental r
166	27	100.0	465	2	B55748	protein kinase (EC	239	27	100.0	673	2	A60692	probable NADH redu
167	27	100.0	467	2	G86461	probable gag-pol p	240	27	100.0	677	1	S09078	chromogranin B pre
168	27	100.0	467	2	T01462	hypothetical prote	241	27	100.0	681	2	T00905	hypothetical prote
169	27	100.0	473	2	JC7359	splicing factor, S	242	27	100.0	682	2	T39613	hypothetical prote
170	27	100.0	474	2	T51103	2,3-dehydratase [v	243	27	100.0	682	2	JC7670	cathepsin B mRNA 3
171	27	100.0	474	2	B70597	hypothetical prote	244	27	100.0	695	2	I51652	dsRNA-binding prot
172	27	100.0	476	2	S57963	methyl CpG binding	245	27	100.0	700	2	I51235	DEAD box protein -
173	27	100.0	477	2	H75267	hypothetical prote	246	27	100.0	707	2	T40070	origin recognition
174	27	100.0	480	2	T44225	hypothetical prote	247	27	100.0	711	2	T47759	hypothetical prote
175	27	100.0	481	2	JE0377	p70 S6 kinase (EC	248	27	100.0	712	2	T18195	gag protein - silk



249	27	100.0	718	2	T02531	probable SCARCEW	322	27	100.0	1256	2	S14556	asparagine-rich pr
250	27	100.0	720	2	T52564	Mre11 protein homo	323	27	100.0	1272	2	T49313	copia-type reverse
251	27	100.0	720	2	T51007	hypothetical prote	324	27	100.0	1299	2	I58401	protein-tyrosine k
252	27	100.0	726	2	D90435	hypothetical prote	325	27	100.0	1307	2	H84467	probable retroelem
253	27	100.0	730	2	E83951	translation initia	326	27	100.0	1320	2	F96614	probable copia-typ
254	27	100.0	722	2	S40759	hypothetical prote	327	27	100.0	1328	2	G84519	probable retroelem
255	27	100.0	744	2	S65869	biotin sulfoxide r	328	27	100.0	1334	2	E86451	probable copia-typ
256	27	100.0	751	2	T48719	hypothetical prote	329	27	100.0	1339	2	T47841	hypothetical prote
257	27	100.0	751	2	T48719	hypothetical prote	330	27	100.0	1352	2	F86246	hypothetical prote
258	27	100.0	755	2	G86469	protein F12K21.6 [	331	27	100.0	1352	2	T47925	copia-type polypro
259	27	100.0	759	2	F86362	F19G10.17 protein	332	27	100.0	1385	2	T21706	hypothetical prote
260	27	100.0	757	2	S41479	DNA-binding protei	333	27	100.0	1411	2	A84460	probable retroelem
261	27	100.0	771	2	S45048	capsid protein - h	334	27	100.0	1422	2	T24212	hypothetical prote
262	27	100.0	772	2	T27512	hypothetical prote	335	27	100.0	1439	2	T02087	gag/pol polyprotei
263	27	100.0	782	2	S33945	late protein, 100K	336	27	100.0	1464	2	T13716	bazooka gene prote
264	27	100.0	796	2	A45695	capsid precursor -	337	27	100.0	1468	2	F96610	probable polyprote
265	27	100.0	800	2	T00034	SART-1 protein - h	338	27	100.0	1490	2	S32373	DNA-binding protei
266	27	100.0	800	2	I51653	dsRNA-binding prot	339	27	100.0	1496	2	T00499	probable retroelem
267	27	100.0	810	2	T44430	protein PV100 [imp	340	27	100.0	1617	2	B86483	protein F5J5.15 [i
268	27	100.0	811	1	F0LJSP	gag polyprotein -	341	27	100.0	1634	2	T28517	hypothetical prote
269	27	100.0	822	2	T51049	related to nucleol	342	27	100.0	1684	2	T02367	hypothetical prote
270	27	100.0	824	2	T20351	hypothetical prote	343	27	100.0	1695	2	T19823	hypothetical prote
271	27	100.0	825	2	JC4163	DNA-binding protei	344	27	100.0	1703	2	S15047	SNF2 protein - yea
272	27	100.0	836	2	G84727	probable DNA topoi	345	27	100.0	1733	1	B45344	probable nuclear a
273	27	100.0	839	2	C85046	probable reverse t	346	27	100.0	1741	2	T13610	parallel sister ch
274	27	100.0	841	2	C87331	ISCC2, transposase	347	27	100.0	1746	1	S19694	tenascin precursor
275	27	100.0	845	2	D96799	hypothetical prote	348	27	100.0	1768	2	T13349	parallel sister ch
276	27	100.0	847	2	S70687	GTP diphosphokinas	349	27	100.0	1846	2	T10670	hypothetical prote
277	27	100.0	848	2	S48273	probable transcript	350	27	100.0	1870	2	S37671	MHC class III hist
278	27	100.0	852	2	B85041	probable receptor	351	27	100.0	1872	2	S36152	MHC class III hist
279	27	100.0	873	2	JC7079	homeobox protein z	352	27	100.0	1858	2	B40505	hypothetical prote
280	27	100.0	873	2	JC4863	hypothetical prote	353	27	100.0	2068	2	A47371	transcription inic
281	27	100.0	879	2	T22033	hypothetical prote	354	27	100.0	2142	2	B35098	MHC class III hist
282	27	100.0	891	2	T26498	hypothetical prote	355	27	100.0	2218	2	B84683	hypothetical prote
283	27	100.0	890	1	F1EC2	translation initia	356	27	100.0	2251	2	T24490	hypothetical prote
284	27	100.0	890	2	D85980	protein chain init	357	27	100.0	2282	2	T42717	DNA-binding protei
285	27	100.0	890	2	A91135	protein chain init	358	27	100.0	2342	2	T13412	hypothetical prote
286	27	100.0	892	2	AH0901	protein chain init	359	27	100.0	2437	2	S53611	MiBP1 protein - ra
287	27	100.0	898	2	H87481	ribonuclease, Rne/	360	27	100.0	2500	1	WMHUE2	HIV-SP2 enhancer-b
288	27	100.0	900	2	A95340	cation transport p	361	27	100.0	2529	2	A56923	transcription fact
289	27	100.0	924	2	E87092	probable ribonucle	362	27	100.0	2561	2	T24864	hypothetical prote
290	27	100.0	925	2	T22388	hypothetical prote	363	27	100.0	2578	2	A56922	transcription fact
291	27	100.0	925	2	A55748	protein kinase (EC	364	27	100.0	2888	2	I49477	alpha-A-crystallin
292	27	100.0	926	2	B84642	hypothetical prote	365	27	100.0	2717	2	A34203	DNA-binding protei
293	27	100.0	933	2	T43201	MutS protein homol	366	27	100.0	2925	2	T00133	RNA-directed RNA p
294	27	100.0	953	2	B70681	probable rne prote	367	27	100.0	3119	2	I49729	HD protein - mouse
295	27	100.0	954	2	G86312	hypothetical prote	368	27	100.0	3566	1	A40701	tenascin-X precurs
296	27	100.0	961	2	T01167	hypothetical prote	369	27	100.0	4006	2	T09070	tenascin-X
297	27	100.0	972	2	T50400	origin recognition	370	27	100.0	4135	2	T42629	tenascin-X - bovin
298	27	100.0	976	2	S57725	respiration defici	371	27	100.0	4957	2	T03455	ALR protein - huma
299	27	100.0	1005	2	B87898	protein F10G8.8 [i	372	27	100.0	5262	2	T03454	ALR protein - huma
300	27	100.0	1027	2	I38604	p53-binding protei	373	24	88.9	41	2	G58208	protamine I - blac
301	27	100.0	1030	2	T16114	hypothetical prote	374	24	88.9	42	2	G95868	hypothetical prote
302	27	100.0	1037	2	T32534	hypothetical prote	375	24	88.9	54	2	G82343	hypothetical prote
303	27	100.0	1041	2	T20730	hypothetical prote	376	24	88.9	93	2	T05020	probable ribosomal
304	27	100.0	1082	2	T42204	chromatin structur	377	24	88.9	94	2	G75143	hypothetical prote
305	27	100.0	1085	2	F96712	hypothetical prote	378	24	88.9	99	2	A55819	nonhistone chromos
306	27	100.0	1088	2	T14917	homeotic protein P	379	24	88.9	101	2	D69333	conserved hypotet
307	27	100.0	1089	2	T36663	protein kinase, tr	380	24	88.9	103	2	T03951	signal recognition
308	27	100.0	1096	2	T21091	hypothetical prote	381	24	88.9	103	2	T46120	hypothetical prote
309	27	100.0	1102	2	A84480	probable retroelem	382	24	88.9	108	2	H85881	hypothetical prote
310	27	100.0	1128	2	G86266	hypothetical prote	383	24	88.9	108	2	F91037	hypothetical prote
311	27	100.0	1139	2	T20660	hypothetical prote	384	24	88.9	108	2	C65013	helix-loop-helix p
312	27	100.0	1148	2	T09073	splicing factor si	385	24	88.9	119	2	S28529	helix-loop-helix p
313	27	100.0	1151	2	T24541	hypothetical prote	386	24	88.9	119	2	A39114	helix-loop-helix p
314	27	100.0	1152	2	H86486	protein Ty./copia-	387	24	88.9	119	2	JC2113	hypothetical prote
315	27	100.0	1173	2	T51892	hypothetical prote	388	24	88.9	127	2	H87298	Na+/H+ antiporter
316	27	100.0	1184	2	A42904	adenylcyclase ty	389	24	88.9	128	2	AB2689	phaf2 protein [imp
317	27	100.0	1188	2	D86311	protein F1L3.20 [i	390	24	88.9	128	2	E97470	conserved hypotet
318	27	100.0	1188	2	G72734	hypothetical prote	391	24	88.9	131	1	B69071	melanoma-derived g
319	27	100.0	1204	2	T19518	hypothetical prote	392	24	88.9	131	2	I38019	DRS12 protein - co
320	27	100.0	1240	2	T04193	hypothetical prote	393	24	88.9	138	2	T12631	hypothetical prote
321	27	100.0	1248	2	T05059	hypothetical prote	394	24	88.9	139	2	H75179	

395	24	88.9	140	2	H71033	hypothetical prote	468	24	88.9	330	2	H70500	probable moex prot
396	24	88.9	141	2	T00928	hypothetical prote	469	24	88.9	331	2	C70465	conserved hypothet
397	24	88.9	141	2	T46015	hypothetical prote	470	24	88.9	332	2	T19150	hypothetical prote
398	24	88.9	145	1	TNLJBT	trans-activating t	471	24	88.9	335	2	D83263	probable hydrolyti
399	24	88.9	149	2	A69948	phage-related prot	472	24	88.9	353	2	E84023	hypothetical prote
400	24	88.9	154	2	T47339	hypothetical prote	473	24	88.9	354	1	A48049	coproporphyrinogen
401	24	88.9	156	2	A72684	hypothetical prote	474	24	88.9	355	1	B69518	GTP-binding protei
402	24	88.9	160	2	S71405	helix-loop-helix p	475	24	88.9	355	2	G84651	biotin holocarboxy
403	24	88.9	162	2	AG1592	B. subtilis PASX p	476	24	88.9	355	2	A33821	chromosomal protei
404	24	88.9	167	2	S73079	hypothetical prote	477	24	88.9	359	2	A13259	hypothetical prote
405	24	88.9	168	2	AE0055	probable lipoprote	478	24	88.9	364	1	A55014	GTP-binding protei
406	24	88.9	169	2	F43368	hypothetical prote	479	24	88.9	364	2	C84187	hypothetical prote
407	24	88.9	174	2	T49558	hypothetical prote	480	24	88.9	366	2	T39192	probable GTP-bind
408	24	88.9	175	1	DBECIB	single-stranded DN	481	24	88.9	366	2	T42381	probable GTP-bind
409	24	88.9	176	2	C69707	RNA polymerase RCF	482	24	88.9	367	1	JC1349	GTP-binding protei
410	24	88.9	177	2	D71869	hypothetical prote	483	24	88.9	367	1	I51426	GTP-binding protei
411	24	88.9	178	2	C38487	single-stranded DN	484	24	88.9	368	1	S42582	GTP-binding protei
412	24	88.9	179	1	DDECF	single-stranded DN	485	24	88.9	368	1	S64487	GTP-binding protei
413	24	88.9	180	1	GCHO	glucagon precursor	486	24	88.9	369	1	S51983	FUN11 protein - ye
414	24	88.9	180	1	GCHY	glucagon precursor	487	24	88.9	369	2	T09368	GTP-binding protei
415	24	88.9	180	1	GCGP	glucagon precursor	488	24	88.9	373	2	C82701	lipoic acid synthe
416	24	88.9	180	1	GCHU	glucagon precursor	489	24	88.9	376	1	MMXREH	nonstructural prot
417	24	88.9	180	1	GCRT	glucagon precursor	490	24	88.9	381	2	T49544	hypothetical prote
418	24	88.9	180	1	GCRTDU	glucagon precursor	491	24	88.9	382	1	B69055	probable cell divi
419	24	88.9	180	2	A57294	glucagon precursor	492	24	88.9	385	2	A75394	conserved hypothet
420	24	88.9	180	2	T23339	glucagon precursor	493	24	88.9	399	2	C96751	probable GTP-bind
421	24	88.9	181	2	B87138	probable adenylate	494	24	88.9	402	2	AC2638	acyl-CoA thiolase
422	24	88.9	181	2	T45390	adenylate kinase l	495	24	88.9	402	2	B57420	probable acyl-CoA
423	24	88.9	181	2	H70822	probable adenylate	496	24	88.9	402	2	T09039	hypothetical prote
424	24	88.9	182	2	T32989	hypothetical prote	497	24	88.9	402	2	D75592	hypothetical prote
425	24	88.9	191	2	PQ0810	glycinin A3B4 (pla	498	24	88.9	404	2	T35428	probable acetyl co
426	24	88.9	198	1	ASLJBT	vif protein - bovi	499	24	88.9	404	2	T35256	probable thiolase
427	24	88.9	209	2	S25617	hypothetical prote	500	24	88.9	407	2	C75218	probable monooxyge
428	24	88.9	211	2	B37499	glial cell line-de	501	24	88.9	414	2	C75461	hypothetical prote
429	24	88.9	211	2	I49686	glial cell line-de	502	24	88.9	417	2	T13759	NADH2 dehydrogenas
430	24	88.9	211	2	A37499	glial cell line-de	503	24	88.9	417	2	AG3571	acetyl-CoA C-acylt
431	24	88.9	217	2	A70774	hypothetical prote	504	24	88.9	419	2	H71117	hypothetical prote
432	24	88.9	218	2	S58769	brain-specific pro	505	24	88.9	420	1	Q0BE44	BDLF2 protein - hu
433	24	88.9	218	2	A96291	hypothetical prote	506	24	88.9	425	2	T48724	hypothetical prote
434	24	88.9	220	2	T09983	hypothetical prote	507	24	88.9	435	2	T22332	hypothetical prote
435	24	88.9	222	2	E83615	probable transcrip	508	24	88.9	461	2	AG3151	glucose 6-phosphat
436	24	88.9	223	2	A83552	probable two-compo	509	24	88.9	461	2	D98136	probable glucose-6
437	24	88.9	227	2	A75204	related biotin ope	510	24	88.9	461	2	AE3208	MFS permease [impo
438	24	88.9	231	2	G86678	hypothetical prote	511	24	88.9	466	2	T18575	cyclophilin isoform
439	24	88.9	233	1	S13625	eosinophil major b	512	24	88.9	469	2	T34173	hypothetical prote
440	24	88.9	233	2	A70351	biotin [acetyl-CoA	513	24	88.9	471	2	S64310	4-aminobutyrate tr
441	24	88.9	234	2	T31676	bact protein - Bac	514	24	88.9	473	2	E72723	probable TARA-bind
442	24	88.9	235	2	B64093	ribosomal protein	515	24	88.9	480	2	S11780	probable transposa
443	24	88.9	236	2	PQ0806	glycinin A3B4 - so	516	24	88.9	484	2	A83082	probable porin PA4
444	24	88.9	237	2	T35855	probable ABC trans	517	24	88.9	490	2	T34448	hypothetical prote
445	24	88.9	238	2	PQ0809	glycinin A3B4 (pla	518	24	88.9	491	1	C8ECAM	UDP-N-acetylmuram
446	24	88.9	243	2	PQ0807	glycinin A3B4 - so	519	24	88.9	491	2	G85491	UDP-N-acetyl-muram
447	24	88.9	248	1	LNDGPS	pulmonary surfacta	520	24	88.9	491	2	G90640	UDP-N-acetyl-muram
448	24	88.9	248	2	A48853	spou rRNA methylas	521	24	88.9	494	2	E71055	hypothetical prote
449	24	88.9	251	2	G87561	hypothetical prote	522	24	88.9	505	2	T05502	hypothetical prote
450	24	88.9	251	2	PQ0808	glycinin A3B4 (pla	523	24	88.9	507	2	F84558	hypothetical prote
451	24	88.9	252	2	D72596	hypothetical prote	524	24	88.9	513	2	T00828	hypothetical prote
452	24	88.9	253	2	T52380	zinc finger protei	525	24	88.9	516	1	FWSY63	glycinin G5 precur
453	24	88.9	253	2	S65170	hypothetical prote	526	24	88.9	521	2	H87111	signal recognition
454	24	88.9	256	1	A57149	gluconate 5-dehydr	527	24	88.9	525	2	D70747	probable fff prote
455	24	88.9	259	2	S75599	hypothetical prote	528	24	88.9	529	2	T00677	hypothetical prote
456	24	88.9	259	2	AC2167	hypothetical prote	529	24	88.9	540	2	S21825	vicilin-like stora
457	24	88.9	273	2	B71355	probable ribosomal	530	24	88.9	543	2	S46713	ATP-dependent RNA
458	24	88.9	274	2	E71257	conserved hypothet	531	24	88.9	545	2	F84533	Mutator-like trans
459	24	88.9	294	2	T31946	hypothetical prote	532	24	88.9	550	2	T34771	signal recognition
460	24	88.9	296	2	G87505	histone deacetylase	533	24	88.9	553	2	JC7723	phytoene desaturas
461	24	88.9	298	2	A13058	gluconate 5-dehydr	534	24	88.9	567	2	H87370	major facilitator
462	24	88.9	298	2	G98227	gluconate 5-dehydr	535	24	88.9	573	2	A53234	globulin-1S, GLB1S
463	24	88.9	305	2	F84434	lectin-like protei	536	24	88.9	573	2	S44605	C02F5.3 protein -
464	24	88.9	310	2	G86490	probable biotin ho	537	24	88.9	592	2	B81009	Bira protein/Bvg a
465	24	88.9	323	2	H89923	hypothetical prote	538	24	88.9	592	2	H82031	probable biotin-la
466	24	88.9	324	2	A99284	hypothetical prote	539	24	88.9	613	2	S19676	68K protein - phag
467	24	88.9	326	1	S76954	probable thioredox	540	24	88.9	621	2	T48646	phytoene dehydroge

541	24	88.9	624	2	T26148	hypothetical prote	614	24	88.9	1264	2	S41603	type V adenyl cy
542	24	88.9	632	2	T25289	hypothetical prote	615	24	88.9	1290	2	T00018	period protein hom
543	24	88.9	639	2	D83591	probable ATP-depen	616	24	88.9	1291	2	T00019	hypothetical prote
544	24	88.9	640	2	T22784	hypothetical prote	617	24	88.9	1308	2	T05178	hypothetical prote
545	24	88.9	644	2	A42220	helix-loop-helix p	618	24	88.9	1314	2	A85176	hypothetical prote
546	24	88.9	644	2	C97005	tetracycline resis	619	24	88.9	1332	2	F69732	PBSX prophage ORF
547	24	88.9	661	2	A81125	bacteriophage tran	620	24	88.9	1415	2	C86438	hypothetical prote
548	24	88.9	671	2	A10816	conserved hypotet	621	24	88.9	1522	2	C96578	hypothetical prote
549	24	88.9	671	2	H91045	hypothetical prote	622	24	88.9	1560	2	T42727	proliferation pote
550	24	88.9	671	2	D85890	hypothetical prote	623	24	88.9	1737	2	A37491	hypothetical helic
551	24	88.9	671	2	A65023	hypothetical prote	624	24	88.9	1840	2	T30250	GPI protein - mous
552	24	88.9	676	2	T47526	protein kinase-lik	625	24	88.9	1879	2	T19481	hypothetical prote
553	24	88.9	682	2	A69170	UDP-N-acetylmuram	626	24	88.9	1977	2	S54771	sodium channel alp
554	24	88.9	692	1	S57592	probable phosphoe	627	24	88.9	2318	2	S45306	notch3 protein - h
555	24	88.9	695	2	I54325	gene XE7 protein -	628	24	88.9	2321	2	S78549	notch3 protein - h
556	24	88.9	699	2	AG0371	probable acetyltra	629	24	88.9	2453	2	S60254	nuclear receptor c
557	24	88.9	708	2	A82377	conserved hypotet	630	24	88.9	2957	2	T33152	hypothetical prote
558	24	88.9	714	2	E83516	hypothetical prote	631	24	88.9	3436	2	S55659	tegument protein 6
559	24	88.9	716	2	C81793	probable ATP-depen	632	23	85.2	48	2	S08608	nonhistone chromos
560	24	88.9	716	2	H81216	hypothetical prote	633	23	85.2	78	2	I51342	homeo box protein
561	24	88.9	716	2	H72635	probable ferric si	634	23	85.2	122	2	T01404	insulin-like growt
562	24	88.9	726	2	AB0122	receptor-like prot	635	23	85.2	122	2	F70510	hypothetical prote
563	24	88.9	732	2	T44483	probable membrane	636	23	85.2	124	2	AE2874	hypothetical prote
564	24	88.9	740	2	F85769	probable iron-sulf	637	23	85.2	136	2	H72633	hypothetical prote
565	24	88.9	740	2	G64919	probable TonB-depe	638	23	85.2	149	2	A70987	hypothetical prote
566	24	88.9	742	2	C83061	conserved hypotet	639	23	85.2	149	2	S68814	legumin beta chain
567	24	88.9	743	2	G90182	conserved hypotet	640	23	85.2	154	2	D87667	conserved hypotet
568	24	88.9	746	2	T46821	siderophore recept	641	23	85.2	155	2	A95989	hypothetical prote
569	24	88.9	746	2	A95420	RhtA Rhizobactin r	642	23	85.2	160	2	D70364	ribosomal protein
570	24	88.9	765	2	A69440	conserved hypotet	643	23	85.2	161	2	C75438	hypothetical prote
571	24	88.9	765	2	D49592	neurofilament trip	644	23	85.2	165	1	I51194	ribosomal protein
572	24	88.9	768	2	D84271	small heat shock p	645	23	85.2	165	1	R3RT10	ribosomal protein
573	24	88.9	770	2	G90506	conserved hypotet	646	23	85.2	165	1	S55918	ribosomal protein
574	24	88.9	772	2	B90921	probable membrane	647	23	85.2	165	2	H59404	ribosomal protein
575	24	88.9	780	1	J43859	ATPase - Sulfolobu	648	23	85.2	165	2	T09252	seed storage prote
576	24	88.9	784	2	JC2055	6-phosphofructokin	649	23	85.2	172	2	T17377	hypothetical prote
577	24	88.9	784	2	G95112	exoribonuclease, V	650	23	85.2	172	2	A75592	hypothetical prote
578	24	88.9	784	2	F97981	potassium channel	651	23	85.2	181	2	AD3277	LSU ribosomal prot
579	24	88.9	787	2	S68699	potassium channel	652	23	85.2	185	2	C69117	conserved hypotet
580	24	88.9	792	2	F83156	probable metal tra	653	23	85.2	185	2	C96808	protein F28K19.4 (
581	24	88.9	816	2	G71127	hypothetical prote	654	23	85.2	191	2	T19573	hypothetical prote
582	24	88.9	817	2	H75035	probable membrane	655	23	85.2	191	2	T22036	hypothetical prote
583	24	88.9	829	2	E64114	translation initia	656	23	85.2	199	2	AB1820	guanylate kinase (
584	24	88.9	834	2	JC8035	N-acetyltransferas	657	23	85.2	207	1	K1ECGU	guanylate kinase (
585	24	88.9	839	2	S32158	DNA topoisomerase	658	23	85.2	207	2	A10969	5'-guanylate kinase
586	24	88.9	840	2	G72468	hypothetical prote	659	23	85.2	207	2	AI0005	guanylate kinase (
587	24	88.9	859	2	C90088	hypothetical prote	660	23	85.2	211	2	C84162	hypothetical prote
588	24	88.9	860	2	AC1250	DNA mismatch repai	661	23	85.2	218	2	AI3612	hypothetical cytos
589	24	88.9	860	2	AG1612	DNA mismatch repai	662	23	85.2	220	2	T18792	hypothetical prote
590	24	88.9	865	2	H71447	trehalose-6-phosph	663	23	85.2	222	1	XUR78C	glutathione trans
591	24	88.9	871	2	H72597	hypothetical prote	664	23	85.2	226	2	E82551	dethiobiotin synth
592	24	88.9	882	2	A42855	N-heparan sulfate	665	23	85.2	227	2	T32894	hypothetical prote
593	24	88.9	882	2	A57169	heparan sulfate-l	666	23	85.2	232	2	B83037	probable ATP-bindi
594	24	88.9	891	2	G84693	probable proline-r	667	23	85.2	233	2	AI3495	probable hydrolase
595	24	88.9	892	2	B46203	mating type A alph	668	23	85.2	238	2	E71258	hypothetical prote
596	24	88.9	940	2	T01854	hypothetical prote	669	23	85.2	241	2	A87409	conserved hypotet
597	24	88.9	943	2	E91286	probable ATP-depen	670	23	85.2	245	2	T15794	hypothetical prote
598	24	88.9	948	2	A57640	retinoblastoma bin	671	23	85.2	247	1	LNRBPS	pulmonary surfacta
599	24	88.9	952	2	A86128	probable ATP-depen	672	23	85.2	251	2	G86368	hypothetical prote
600	24	88.9	985	2	T29910	hypothetical prote	673	23	85.2	251	2	G86368	enoyl-CoA hydratase
601	24	88.9	1006	2	T41104	probable transcrip	674	23	85.2	256	2	F87668	D-beta-hydroxybuty
602	24	88.9	1024	2	T46016	hypothetical prote	675	23	85.2	265	2	F87668	K+ channel-interac
603	24	88.9	1033	2	T38131	hypothetical prote	676	23	85.2	270	2	JC7631	hypothetical prote
604	24	88.9	1043	2	T15191	hypothetical prote	677	23	85.2	272	2	C42384	hypothetical prote
605	24	88.9	1052	2	T00067	hypothetical prote	678	23	85.2	278	2	T45494	probable transposa
606	24	88.9	1056	2	S55151	probable membrane,	679	23	85.2	281	2	B86356	T16E15.6 protein -
607	24	88.9	1095	2	T43275	neurabin - rat	680	23	85.2	298	2	C75444	hypothetical prote
608	24	88.9	1128	2	AD3008	peptide synthetase	681	23	85.2	300	2	T49595	conserved hypotet
609	24	88.9	1137	2	A98276	hypothetical prote	682	23	85.2	302	2	F81016	conserved hypotet
610	24	88.9	1154	2	S43275	hypothetical prote	683	23	85.2	302	2	H96811	methionyl-tRNA for
611	24	88.9	1222	2	G72614	probable reverse g	684	23	85.2	308	2	F87283	plastoquinol-plast
612	24	88.9	1231	2	T24415	hypothetical prote	685	23	85.2	314	2	S78363	plastoquinol-plast
613	24	88.9	1236	2	T50904	Mg protoporphyrin	686	23	85.2	317	2	S16916	plastoquinol-plast
										318	2	S00431	

687	23	85.2	319	2	T14833	plastoquinol-plast	760	23	85.2	487	2	F84751	hypothetical prote
688	23	85.2	319	2	T07480	plastoquinol-plast	761	23	85.2	492	1	PWLRE	l2 protein - cotto
689	23	85.2	320	1	CFLV	plastoquinol-plast	762	23	85.2	497	2	E81987	hypothetical prote
690	23	85.2	320	1	CFNT	plastoquinol-plast	763	23	85.2	498	2	S11246	LAG-3 protein prec
691	23	85.2	320	1	CFRZ	plastoquinol-plast	764	23	85.2	509	2	F81041	cytoplasmic axial
692	23	85.2	320	2	S73186	cytochrome f - red	765	23	85.2	525	2	JN0059	hypothetical 57.4K
693	23	85.2	320	2	S07296	plastoquinol-plast	766	23	85.2	537	2	T48599	hypothetical prote
694	23	85.2	320	2	S85664	plastoquinol-plast	767	23	85.2	539	2	T50435	conserved hypothet
695	23	85.2	320	2	S04330	plastoquinol-plast	768	23	85.2	544	2	S43058	CCTeta protein eta
696	23	85.2	320	2	S45661	plastoquinol-plast	769	23	85.2	550	1	S14048	RNA helicase dbp2
697	23	85.2	320	2	T06347	plastoquinol-plast	770	23	85.2	552	2	T33511	hypothetical prote
698	23	85.2	320	2	A26576	plastoquinol-plast	771	23	85.2	558	1	A53719	glutamate dehydrog
699	23	85.2	321	2	T06932	plastoquinol-plast	772	23	85.2	560	2	S23313	hypothetical prote
700	23	85.2	328	2	S18573	plastoquinol-plast	773	23	85.2	562	2	B70609	hypothetical prote
701	23	85.2	331	2	T44902	plastoquinol-plast	774	23	85.2	567	2	T43555	Ras pathway intera
702	23	85.2	333	1	I52525	probable riboflavi	775	23	85.2	569	2	T35476	probable regulator
703	23	85.2	333	2	B35580	plastoquinol-plast	776	23	85.2	573	2	A86287	hypothetical prote
704	23	85.2	333	2	AE2112	apocytochrome f li	777	23	85.2	581	2	S46742	hypothetical prote
705	23	85.2	337	2	AF0466	probable lysophosp	778	23	85.2	582	2	T05020	L-ascorbate oxidas
706	23	85.2	338	2	AE0917	lysophospholipase	779	23	85.2	586	2	C83262	hypothetical prote
707	23	85.2	338	2	F82358	lysophospholipase	780	23	85.2	613	2	T24662	hypothetical prote
708	23	85.2	338	2	S04045	embryonic abundant	781	23	85.2	615	2	F87599	hypothetical prote
709	23	85.2	340	1	PSECL2	lysophospholipase	782	23	85.2	623	2	T07933	polyadenylate-bind
710	23	85.2	340	2	A86070	lysophospholipase	783	23	85.2	631	2	T40189	probable biotin-pr
711	23	85.2	340	2	C91223	lysophospholipase	784	23	85.2	638	2	T24661	hypothetical prote
712	23	85.2	341	2	T07339	plastoquinol-plast	785	23	85.2	653	2	A46362	amyloid precursor-
713	23	85.2	341	2	T08459	hypothetical prote	786	23	85.2	688	2	A49318	protein kinase (BC
714	23	85.2	342	1	T48847	syntaxin synt4 fim	787	23	85.2	691	2	F91251	probable tape meas
715	23	85.2	342	1	CFPM	plastoquinol-plast	788	23	85.2	692	2	T13161	A-kinase anchor pr
716	23	85.2	344	2	S61037	hypothetical prote	789	23	85.2	701	2	T20892	hypothetical prote
717	23	85.2	345	2	H71358	conserved hypothet	790	23	85.2	716	2	T03695	delta 1 pyroline-
718	23	85.2	347	2	E75251	probable cytochrom	791	23	85.2	761	2	A53414	A-kinase anchor pr
719	23	85.2	349	2	C81007	probable ATP-bind	792	23	85.2	769	2	F87486	outer membrane pro
720	23	85.2	352	2	JC5388	replication initia	793	23	85.2	770	2	D87315	conserved hypothet
721	23	85.2	359	2	T46485	hypothetical prote	794	23	85.2	771	2	S18624	ntry protein - Azo
722	23	85.2	360	2	S48365	hypothetical prote	795	23	85.2	793	2	G84475	hypothetical prote
723	23	85.2	361	2	T13745	hypothetical prote	796	23	85.2	829	2	S58888	Ins P4-binding pro
724	23	85.2	367	2	I40843	heat shock protein	797	23	85.2	829	2	S71847	Ins P4-binding pro
725	23	85.2	367	2	S16562	nolr protein - Rhi	798	23	85.2	859	2	T01672	envelope polyprote
726	23	85.2	369	2	F70905	probable lipoprote	799	23	85.2	886	2	S07132	hypothetical prote
727	23	85.2	374	2	AF0304	probable dioxygena	800	23	85.2	956	2	H75536	protein T19E23.7 l
728	23	85.2	374	2	C85791	probable choline m	801	23	85.2	1014	2	H86438	2-oxoglutarate deh
729	23	85.2	374	2	G90942	probable choline m	802	23	85.2	1022	2	T24663	hypothetical prote
730	23	85.2	374	2	B64941	probable choline m	803	23	85.2	1026	2	G87346	hypothetical prote
731	23	85.2	377	2	T47439	hypothetical prote	804	23	85.2	1063	1	GNWV77	structural polypro
732	23	85.2	390	2	T50880	hydroxyneurosporen	805	23	85.2	1063	1	GNWVR4	structural polypro
733	23	85.2	392	2	B86495	heat shock protein	806	23	85.2	1063	1	GNWVR4	structural polypro
734	23	85.2	392	2	G72128	heat shock protein	807	23	85.2	1078	2	T30860	tra1 protein - Sal
735	23	85.2	392	2	D81683	dnal protein TC061	808	23	85.2	1119	2	A88481	protein C16A3.6 li
736	23	85.2	392	2	H71526	probable heat choc	809	23	85.2	1134	2	G89269	molybdopterin-bind
737	23	85.2	396	1	KX80Z	plasma protein Z -	810	23	85.2	1140	2	D70729	hypothetical prote
738	23	85.2	444	2	AH0346	putative HlyD fami	811	23	85.2	1165	2	D72496	hypothetical prote
739	23	85.2	448	2	A82679	phosphomannomutase	812	23	85.2	1168	1	MWAXIC	myosin heavy chain
740	23	85.2	448	2	G95038	IS1380-Spn1, trans	813	23	85.2	1199	2	T15826	hypothetical prote
741	23	85.2	448	2	D95040	IS1380-Spn1, trans	814	23	85.2	1291	1	S05465	retrovirus-related
742	23	85.2	448	2	D95057	IS1380-Spn1, trans	815	23	85.2	1344	2	T14316	rig-1 protein - mo
743	23	85.2	448	2	F95082	IS1380-Spn1, trans	816	23	85.2	1366	2	T35985	probable large pro
744	23	85.2	448	2	B95155	IS1380-Spn1, trans	817	23	85.2	1402	2	D70634	probable polyketid
745	23	85.2	448	2	A95157	IS1380-Spn1, trans	818	23	85.2	1413	2	D84481	probable retroelem
746	23	85.2	448	2	B95165	IS1380-Spn1, trans	819	23	85.2	1797	2	A55677	laminin beta-2 cha
747	23	85.2	448	2	F95167	IS1380-Spn1, trans	820	23	85.2	1798	2	S53869	laminin beta-2 cha
748	23	85.2	448	2	F95175	IS1380-Spn1, trans	821	23	85.2	1942	2	B71426	hypothetical prote
749	23	85.2	448	2	G95254	IS1380-Spn1, trans	822	23	85.2	2265	1	FNBU	fibronectin - bovi
750	23	85.2	448	2	G95254	IS1380-Spn1, trans	823	23	85.2	2386	1	FNBU	fibronectin precur
751	23	85.2	450	2	A34169	alpha-2A-adrenergi	824	23	85.2	2477	2	S14428	fibronectin - Afri
752	23	85.2	450	2	AG2662	GDH family prote	825	23	85.2	2481	2	A43908	fibronectin - Afri
753	23	85.2	450	2	F97444	hypothetical prote	826	23	85.2	2605	2	T18552	saframycin Mxi syn
754	23	85.2	451	2	AD1811	UDP-N-acetylmurano	827	23	85.2	3172	2	S22012	erythronolide synt
755	23	85.2	454	2	S75610	UDP-N-acetylmurano	828	23	85.2	3178	2	S13595	6-deoxyerythronoli
756	23	85.2	455	2	A10599	probable ATP-depen	829	23	85.2	3591	1	S21010	filamentous hemagg
757	23	85.2	463	2	E83468	probable sodium/so	830	23	85.2	4544	1	S02392	alpha-2-macroglobu
758	23	85.2	468	2	D70532	hypothetical prote	831	22	81.5	17	2	PH1357	IG heavy chain DJ
759	23	85.2	476	2	A82555	conserved hypothet	832	22	81.5	18	2	B24735	glutathione transf

833	22	81.5	20	2	E60894	crystallin - Pacif	906	22	81.5	90	2	A49140	glutathione transf
834	22	81.5	26	2	A61056	aminopyrine N-deme	907	22	81.5	90	2	T36074	hypothetical prote
835	22	81.5	28	2	S21378	glutathione trans	908	22	81.5	91	2	JT0759	gag-like protein -
836	22	81.5	28	2	S29135	aminopyrine N-deme	909	22	81.5	91	2	C72638	probable ribosomal
837	22	81.5	28	2	S29136	aminopyrine N-deme	910	22	81.5	91	2	T36618	hypothetical prote
838	22	81.5	32	2	D32502	T-cell receptor de	911	22	81.5	93	2	E87678	hypothetical prote
839	22	81.5	32	2	PQ0425	nonstructural prot	912	22	81.5	94	1	H75022	ribosomal protein
840	22	81.5	32	2	PQ0413	nonstructural prot	913	22	81.5	94	2	H83655	probable insertion
841	22	81.5	35	2	A61375	basic fibroblast g	914	22	81.5	94	2	T36982	hypothetical prote
842	22	81.5	37	2	S29829	protamine Z3 - sma	915	22	81.5	95	2	T45263	cobalt transport p
843	22	81.5	38	2	A05149	crystallin SIII -	916	22	81.5	95	2	T14743	hypothetical prote
844	22	81.5	41	2	A58213	protamine - green	917	22	81.5	96	2	B37470	UL49.5 homolog - b
845	22	81.5	45	2	T07502	hypothetical prote	918	22	81.5	96	2	A81031	hypothetical prote
846	22	81.5	47	2	JT0518	Ig heavy chain V-I	919	22	81.5	97	2	T36615	hypothetical prote
847	22	81.5	47	2	S55777	MADS box protein Z	920	22	81.5	99	2	S20765	Ig heavy chain V r
848	22	81.5	48	2	S20256	omega-agatoxin IVA	921	22	81.5	99	2	B87374	transcription regu
849	22	81.5	52	2	C91112	hypothetical prote	922	22	81.5	99	2	T37093	hypothetical prote
850	22	81.5	54	2	S34567	gene L protein - h	923	22	81.5	101	2	T37064	hypothetical prote
851	22	81.5	56	2	S49274	ribulose-bisphosph	924	22	81.5	101	2	C82133	hypothetical prote
852	22	81.5	56	2	S49275	ribulose-bisphosph	925	22	81.5	102	2	T49343	hypothetical prote
853	22	81.5	56	2	S49276	ribulose-bisphosph	926	22	81.5	102	2	AD2670	hypothetical prote
854	22	81.5	56	2	S49277	ribulose-bisphosph	927	22	81.5	102	2	B97452	hypothetical prote
855	22	81.5	56	2	S49278	ribulose-bisphosph	928	22	81.5	102	2	D71415	hypothetical prote
856	22	81.5	56	2	S49281	ribulose-bisphosph	929	22	81.5	103	2	AF3455	hypothetical prote
857	22	81.5	56	2	S49282	ribulose-bisphosph	930	22	81.5	104	2	A24424	T-cell receptor be
858	22	81.5	56	2	S49283	ribulose-bisphosph	931	22	81.5	104	2	T24609	hypothetical prote
859	22	81.5	56	2	S49284	ribulose-bisphosph	932	22	81.5	105	2	G86927	conserved hypotet
860	22	81.5	56	2	S49285	ribulose-bisphosph	933	22	81.5	105	2	B70716	hypothetical prote
861	22	81.5	56	2	S49286	ribulose-bisphosph	934	22	81.5	105	2	T38300	ribosomal protein
862	22	81.5	56	2	S49297	ribulose-bisphosph	935	22	81.5	105	2	E69058	hypothetical prote
863	22	81.5	56	2	S49298	ribulose-bisphosph	936	22	81.5	105	2	H72533	hypothetical prote
864	22	81.5	56	2	T34739	ribulose-bisphosph	937	22	81.5	105	2	A43861	vir-repressed prot
865	22	81.5	57	2	PL0225	hydrophobic protei	938	22	81.5	105	2	T00283	hypothetical prote
866	22	81.5	57	2	S34045	T-cell receptor be	939	22	81.5	106	2	E75371	hypothetical prote
867	22	81.5	58	2	T04616	protamine - North	940	22	81.5	106	2	D87618	hypothetical prote
868	22	81.5	59	2	PL0166	hypothetical prote	941	22	81.5	106	2	B64350	hypothetical prote
869	22	81.5	59	2	D41476	amyloid protein AA	942	22	81.5	107	2	S02458	hypothetical prote
870	22	81.5	60	2	JT0512	probable antigen 4	943	22	81.5	107	2	S07818	glutathione transf
871	22	81.5	60	2	H70809	Ig heavy chain V-I	944	22	81.5	108	1	B3AG55	homeotic protein H
872	22	81.5	61	2	S55773	hypothetical prote	945	22	81.5	108	1	B3AG58	virB3 protein - Ag
873	22	81.5	61	2	A82099	homeotic protein Z	946	22	81.5	108	2	AG3248	component of type
874	22	81.5	62	2	G87184	hypothetical prote	947	22	81.5	108	2	T47985	hypothetical prote
875	22	81.5	63	2	JC7600	conserved hypotet	948	22	81.5	109	2	A49009	DNA-binding transc
876	22	81.5	64	2	S01103	vesicular membrane	949	22	81.5	109	2	F85135	hypothetical prote
877	22	81.5	65	2	T13352	hypothetical prote	950	22	81.5	109	2	T03257	glycine-rich prote
878	22	81.5	66	2	T35419	hypothetical prote	951	22	81.5	109	2	E72527	hypothetical prote
879	22	81.5	66	2	A64350	small hypothetical	952	22	81.5	110	2	T26729	hypothetical prote
880	22	81.5	66	2	H69548	hypothetical prote	953	22	81.5	111	2	S76083	hypothetical prote
881	22	81.5	68	2	AD2807	hypothetical prote	954	22	81.5	111	2	B72722	hypothetical prote
882	22	81.5	70	4	I52626	hypothetical FUS/E	955	22	81.5	112	2	S14355	glutathione transf
883	22	81.5	72	2	S56770	RNA-directed RNA p	956	22	81.5	112	2	B43601	LSR2 T-cell antige
884	22	81.5	76	2	C82511	hypothetical prote	957	22	81.5	112	2	F70954	probable lsr2 prot
885	22	81.5	76	2	G75602	hypothetical prote	958	22	81.5	113	2	S26263	T-cell receptor be
886	22	81.5	76	2	E97763	protein transport	959	22	81.5	113	2	S26262	T-cell receptor be
887	22	81.5	78	2	C95976	hypothetical prote	960	22	81.5	113	2	S17385	T-cell receptor be
888	22	81.5	81	2	E97791	hypothetical prote	961	22	81.5	113	2	C70691	hypothetical prote
889	22	81.5	82	2	H82586	hypothetical prote	962	22	81.5	113	2	A90878	hypothetical prote
890	22	81.5	83	2	S77417	prochlorophyllid	963	22	81.5	113	2	AD0582	probable hydrolase
891	22	81.5	83	2	T35979	hypothetical prote	964	22	81.5	113	2	F95919	probable transposa
892	22	81.5	84	2	B72065	ribosomal protein	965	22	81.5	114	2	T18108	hypothetical prote
893	22	81.5	84	2	E86558	L27 ribosomal prot	966	22	81.5	115	2	JC5908	T cell receptor va
894	22	81.5	85	2	F86389	probable DNA-bindi	967	22	81.5	115	2	T37901	histone H2A relate
895	22	81.5	85	2	A30924	hypothetical 10.8k	968	22	81.5	115	2	T36493	probable gntR-fami
896	22	81.5	85	2	A13150	hypothetical prote	969	22	81.5	115	2	T06772	early noduline 7 -
897	22	81.5	86	2	C90666	probable DNA bindi	970	22	81.5	115	2	D95310	hypothetical prote
898	22	81.5	86	2	G85516	probable DNA bindi	971	22	81.5	115	2	AD2303	flex protein [impor
899	22	81.5	86	2	T20282	hypothetical prote	972	22	81.5	116	2	T36999	probable transposa
900	22	81.5	87	2	I68792	MHC class II histo	973	22	81.5	116	2	H75601	hypothetical prote
901	22	81.5	87	2	A55846	oppF homolog, kmb	974	22	81.5	117	2	S09961	Ig heavy chain V-D
902	22	81.5	87	2	E75417	hypothetical prote	975	22	81.5	118	2	AB2862	conserved hypotet
903	22	81.5	87	2	G70663	hypothetical prote	976	22	81.5	118	2	A97639	hypothetical prote
904	22	81.5	87	2	E81393	probable altronate	977	22	81.5	118	2	I55064	transposase - Esch
905	22	81.5	89	2	B92982	nonstructural prot	978	22	81.5	118	2	C72642	hypothetical prote

```
979      22      81.5      118      2      S26611      transcription fact
980      22      81.5      119      2      A64249      ribosomal protein
981      22      81.5      120      2      JC2475      growth hormone reg
982      22      81.5      121      2      D86468      hypothetical prote
983      22      81.5      122      1      GCAF2      glucagon 2 precurs
984      22      81.5      122      2      C30248      serum amyloid AA2
985      22      81.5      122      2      B75177      hypothetical prote
986      22      81.5      122      2      S72797      probable K+ transp
987      22      81.5      122      2      AG0734      probable bacteriop
988      22      81.5      122      2      A86817      hypothetical prote
989      22      81.5      122      2      T71037      hypothetical prote
990      22      81.5      123      2      T16234      hypothetical prote
991      22      81.5      123      2      B69050      conserved hypothet
992      22      81.5      123      2      S14573      globin B precursor
993      22      81.5      123      2      H72709      hypothetical prote
994      22      81.5      123      2      D72579      hypothetical prote
995      22      81.5      124      1      S15625      E4 protein - human
996      22      81.5      124      2      F69023      conserved hypothet
997      22      81.5      124      2      T49423      hypothetical prote
998      22      81.5      125      1      A46315      E4 protein - human
999      22      81.5      125      2      S53086      NADH2 dehydrogenas
1000     22      81.5      125      2      T49356      hypothetical prote

ALIGNMENTS

RESULT 1
T36132
hypothetical protein SCE19A.09 - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T36132
R:Seeger, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1999
A:Reference number: Z21598
A:Accession: T36132
A:Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: DNA
A:Residues: 1-55 <SEE>
A:Cross-references: UNIPROT:Q9S2H2; UNIPARC:UPI000000DB270; EMBL:AL096852; PIDN:CAB50990.
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SCE19A.09

Query Match          100.0%; Score 27; DB 2; Length 55;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1      RGRGR 5
Db      32      RGRGR 36

RESULT 2
I54837
Smd cross-reacting protein - mouse (fragment)
C:Species: Mus musculus (house mouse)
C>Date: 27-Feb-1997 #sequence_revision 27-Feb-1997 #text_change 05-Nov-1999
C:Accession: I54837
R:Rivkin, E.; Vella, M.J.; Lahita, R.G.
J. Autoimmun. 7, 119-132, 1994
A:Title: A heterogeneous immune response to an Smd-like epitope by SLE patients.
A:Reference number: I54837; MUID:94257110; PMID:7515243
A:Accession: I54837
A:Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: mRNA
A:Residues: 1-59 <RES>
A:Cross-references: UNIPARC:UPI000000E7233; GB:S71494; NID:9557919; PIDN:AAB30912.1; PID:

Query Match          100.0%; Score 27; DB 2; Length 59;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1      RGRGR 5
Db      17      RGRGR 21

RESULT 3
S39424
protamine P1 - Australian echidna
C:Species: Tachyglossus aculeatus (Australian echidna)
C>Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 09-Jul-2004
C:Accession: S39424
R:Kretief, J.D.; Winkfein, R.J.; Dixon, G.H.
Eur. J. Biochem. 218, 457-461, 1993
A:Title: Evolution of the monotremes. The sequences of the protamine P1 genes of platypus
A:Reference number: S39424; MUID:94094837; PMID:8269934
A:Accession: S39424
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-69 <RET>
A:Cross-references: UNIPROT:P35311; UNIPARC:UPI000016C71A; EMBL:Z26848; NID:G407183; PID:
C:Genetics:
A:Introns: 53/1
C:Superfamily: sperm histone
C:Keywords: DNA binding

Query Match          100.0%; Score 27; DB 2; Length 69;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1      RGRGR 5
Db      41      RGRGR 45

RESULT 4
JC5931
high mobility group I HMGI chromosomal protein isoform C-alpha - human
C:Species: Homo sapiens (man)
C>Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 07-May-1999
C:Accession: JC5931
R:Kottickal, L.V.; Sarada, B.; Ashar, H.; Chada, K.; Nagarajan, L.
Biochem. Biophys. Res. Commun. 242, 452-456, 1998
A:Title: Preferential expression of HMGI-C isoforms lacking the acidic carboxy terminal i
A:Reference number: JC5931; MUID:98113374; PMID:9446816
A:Accession: JC5931
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-90 <NOT>
A:Cross-references: UNIPARC:UPI00001771E6
C:Superfamily: nonhistone chromosomal protein HMG-I

Query Match          100.0%; Score 27; DB 2; Length 90;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1      RGRGR 5
Db      27      RGRGR 31

RESULT 5
H72271
ferredoxin - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 31-Dec-2004
C:Accession: H72271
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey,
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316; PMID:10360571
```

```

A;Accession: H72271
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-95 <ARN>
A;Cross-references: UNIPROT:Q9X115; UNIPARC:UPI00000D734E; GB:AE001784; GB:AE000512; NID
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TW1289
C;Superfamily: pyruvate synthase, PorD subunit; ferredoxin 2[4Fe-4S] homology
C;Keywords: 4Fe-4S; iron-sulfur protein; metalloprotein
F;8,11,14,48/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted
F;18,38,41,44/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 100.0%; Score 27; DB 2; Length 95;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
Db 87 RGRGR 91

RESULT 6
JC5933
high mobility group I HMGI chromosomal protein isoform C-gamma - human
C;Species: Homo sapiens (man)
C;Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 07-May-1999
C;Accession: JC5933
R;Kottickal, L.V.; Sarada, B.; Ashar, H.; Chada, K.; Nagarajan, L.
Biochem. Biophys. Res. Commun. 242, 452-456, 1998
A;Title: Preferential expression of HMGI-C isoforms lacking the acidic carboxy terminal
A;Reference number: JC5931; MUID:98113374; PMID:9446816
A;Accession: JC5933
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-96 <KOT>
A;Cross-references: UNIPARC:UPI00001771E8
C;Superfamily: nonhistone chromosomal protein HMGI-I

Query Match 100.0%; Score 27; DB 2; Length 96;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
Db 27 RGRGR 31

RESULT 7
A31895
nonhistone chromosomal protein HMGI-I(Y) - mouse
C;Species: Mus musculus (house mouse)
C;Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 09-Jul-2004
C;Accession: A31895
R;Johnson, K.R.; Lehn, D.A.; Elton, T.S.; Barr, P.J.; Reeves, R.
J. Biol. Chem. 263, 18338-18342, 1988
A;Title: Complete murine cDNA sequence, genomic structure, and tissue expression of the
A;Reference number: A31895; MUID:89054013; PMID:3192537
A;Accession: A31895
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-96 <JOH>
A;Cross-references: UNIPROT:P17095; UNIPARC:UPI00000018BD; GB:J04179; NID:g193883; PIDN:
C;Superfamily: nonhistone chromosomal protein HMGI-I
C;Keywords: alternative splicing; chromosomal protein; DNA binding; nucleus; phosphoprotein

Query Match 100.0%; Score 27; DB 2; Length 96;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
Db 24 RGRGR 28

```

## RESULT 8

```

D97652
hypothetical protein AGR_C_4419 [imported] - Agrobacterium tumefaciens (strain C58, Cerc
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
C;Accession: D97652
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Outololo, B.; Goldman,
A.; Liu, F.; Wollan, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: D97652
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-99 <KUR>
A;Cross-references: UNIPROT:Q8U540; UNIPARC:UPI00000D28E8; GB:AE007869; PIDN:AAK88173.1;
C;Genetics:
A;Gene: AGR_C_4419
A;Map position: circular chromosome

```

```

Query Match 100.0%; Score 27; DB 2; Length 99;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 RGRGR 5
Db 29 RGRGR 33

```

## RESULT 9

```

T44039
replication protein [imported] - human herpesvirus 6 (strain HST)
C;Species: human herpesvirus 6
A;Variety: strain HST
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
C;Accession: T44039
R;Isegawa, Y.; Mukai, T.; Nakano, K.; Kagawa, M.; Chen, J.; Mori, Y.; Sunagawa, T.; Kawa
J. Virol. 73, 8053-8063, 1999
A;Title: Comparison of the complete DNA sequences of human herpesvirus 6 variants A and
A;Reference number: Z22732; MUID:99412319; PMID:10482554
A;Accession: T44039
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-100 <ISE>
A;Cross-references: UNIPROT:Q96896; UNIPARC:UPI00000F0D92; EMBL:AB021506; NID:g4995977; I
A;Experimental source: strain HST; pop. variant B
C;Genetics:
A;Note: U80

```

```

Query Match 100.0%; Score 27; DB 2; Length 100;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 RGRGR 5
Db 45 RGRGR 49

```

## RESULT 10

```

JC5932
high mobility group I HMGI chromosomal protein isoform C-beta - human
C;Species: Homo sapiens (man)
C;Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 07-May-1999
C;Accession: JC5932
R;Kottickal, L.V.; Sarada, B.; Ashar, H.; Chada, K.; Nagarajan, L.
Biochem. Biophys. Res. Commun. 242, 452-456, 1998
A;Title: Preferential expression of HMGI-C isoforms lacking the acidic carboxy terminal
A;Reference number: JC5931; MUID:98113374; PMID:9446816
A;Accession: JC5932
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-105 <KOT>

```

A;Cross-references: UNIPARC:UPI00001771E7  
 C;Superfamily: nonhistone chromosomal protein HMG-I

Query Match 100.0%; Score 27; DB 2; Length 105;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
 |||||  
 Db 27 RGRGR 31

## RESULT 11

T49594  
 nonhistone chromosomal protein HMG-I - human  
 N:Alternate names: high-mobility group protein HMG-I; HMG1Y splice form I  
 N:Contains: nonhistone chromosomal protein HMG-Y  
 C:Species: Homo sapiens (man)  
 C>Date: 08-Dec-1989 #sequence revision 08-Dec-1989 #text change 09-Jul-2004  
 C:Accession: A32794; B32794; S05322; S41769; A26907; A32250  
 R:Johnson, K.R.; Lehn, D.A.; Reeves, R.  
 Mol. Cell. Biol. 9, 2114-2123, 1989  
 A>Title: Alternative processing of mRNAs encoding mammalian chromosomal high-mobility-group proteins  
 A:Reference number: A32794; MUID:89313758; PMID:2701943  
 A:Accession: A32794  
 A:Molecule type: mRNA  
 A:Residues: 1-107 <JOH>  
 A:Cross-references: UNIPROT:P17096; UNIPARC:UPI0000000C96; GB:M23614; NID:gl84246; PIDN:A32794  
 A:Accession: B32794  
 A:Molecule type: mRNA  
 A:Residues: 1-34,46-107 <JO2>  
 A:Cross-references: UNIPARC:UPI0000000C66; GB:M23615; NID:gl84252; PIDN:AAA88073.1; PIDN:R:Eckner, R.; Birnstiel, M.L.  
 Nucleic Acids Res. 17, 5947-5959, 1989  
 A>Title: Cloning of cDNAs coding for human HMG I and HMG Y proteins: both are capable of DNA binding  
 A:Reference number: S05321; MUID:89366631; PMID:2505228  
 A:Accession: S05322  
 A:Molecule type: mRNA  
 A:Residues: 1-107 <ECK>  
 A:Cross-references: UNIPARC:UPI0000000C96; EMBL:X14957; NID:g32336; PIDN:CAA33080.1; PIDN:R:Friedmann, M.; Holth, L.T.; Zoghbi, H.Y.; Reeves, R.  
 Nucleic Acids Res. 21, 4259-4267, 1993  
 A>Title: Organization, inducible-expression and chromosome localization of the human HMG Y gene  
 A:Reference number: S41769; MUID:94021372; PMID:8414980  
 A:Accession: S41769  
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-107 <FRI>  
 A:Cross-references: UNIPARC:UPI0000000C96; EMBL:L17131; NID:g306868; PIDN:AA800145.1; PIDN:R:Lund, T.; Dahl, K.H.; Mork, E.; Holtlund, J.; Laland, S.G.  
 Biochem. Biophys. Res. Commun. 146, 725-730, 1987  
 A>Title: The human chromosomal protein HMG I contains two identical palindromic amino acid repeats  
 A:Reference number: A26907; MUID:87298501; PMID:3619901  
 A:Accession: A26907  
 A:Molecule type: protein  
 A:Residues: 4-25, 'X' 27-68, 69-107, 'O' <LUN>  
 A:Cross-references: UNIPARC:UPI00001771E3; UNIPARC:UPI00001771E4  
 R:Carlson, J.R.; Mork, E.; Holtlund, J.; Laland, S.G.; Lund, T.  
 Biochem. Biophys. Res. Commun. 158, 646-651, 1989  
 A>Title: The amino acid sequence of the chromosomal protein HMG-Y, its relation to HMG-I and HMG-X  
 A:Reference number: A32250; MUID:89149775; PMID:2920035  
 A:Accession: A32250  
 A:Molecule type: protein  
 A:Residues: 4-34,46-107, 'QQ' <KAR>  
 A:Cross-references: UNIPARC:UPI00001771E5  
 C:Genetics:  
 A:Gene: GDB:HMGIY  
 A:Cross-references: GDB:134205; OMIM:600701  
 A:Map position: 6p21-6p21  
 A:Introns: 45/3; 73/3; 90/3  
 C;Superfamily: nonhistone chromosomal protein HMG-I  
 C;Keywords: alternative splicing; blocked amino end; chromosomal protein; DNA binding; n

F:1-107/Product: nonhistone chromosomal protein HMG-I #status predicted <MATI>  
 F:1-34,46-107/Product: nonhistone chromosomal protein HMG-Y #status predicted <MATY>  
 F:102,103/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 100.0%; Score 27; DB 2; Length 107;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
 |||||  
 Db 24 RGRGR 28

## RESULT 12

T49594  
 hypothetical protein B3E4.70 [imported] - Neurospora crassa  
 C:Species: Neurospora crassa  
 C>Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 18-Aug-2000  
 C:Accession: T49594  
 R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura, submitted to the Protein Sequence Database, May 2000  
 A:Reference number: Z25022  
 A:Accession: T49594  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-107 <SCH>  
 A:Cross-references: UNIPARC:UPI0000179D4E; EMBL:AL355931; GSPDB:GN00116; NCSP:B3E4.70  
 A:Experimental source: BAC clone B3E4; strain OR74A  
 C:Genetics:  
 A:Gene: NCSP:B3E4.70  
 A:Map position: 6  
 A:Introns: 28/3  
 C:Superfamily: Neurospora crassa hypothetical protein B3E4.70

Query Match 100.0%; Score 27; DB 2; Length 107;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
 |||||  
 Db 26 RGRGR 30

## RESULT 13

JC4575  
 high-mobility-group I protein C - mouse  
 N:Alternate names: nuclear phosphoprotein  
 C:Species: Mus musculus (house mouse)  
 C>Date: 10-Apr-1996 #sequence\_revision 24-May-1996 #text\_change 09-Jul-2004  
 C:Accession: JC4575; S22597  
 R:Manfioletti, G.; Rustighi, A.; Mantovani, F.; Goodwin, G.H.; Giancotti, V.  
 Gene 167, 249-253, 1995  
 A>Title: Isolation and characterization of the gene coding for murine high-mobility-group I protein  
 A:Reference number: JC4575; MUID:96144283; PMID:8566786  
 A:Accession: JC4575  
 A:Molecule type: mRNA  
 A:Residues: 1-108 <MAN>  
 A:Cross-references: UNIPROT:P52927; UNIPARC:UPI0000003F65; GB:L41617  
 R:Manfioletti, G.; Giancotti, V.; Bandiera, A.; Buratti, E.; Sautiere, P.; Cary, P.; Crai  
 Nucleic Acids Res. 19, 6793-6797, 1991  
 A>Title: cDNA cloning of the HMG1-C phosphoprotein, a nuclear protein associated with ne  
 A:Reference number: S22597; MUID:92107664; PMID:1762909  
 A:Accession: S22597  
 A:Molecule type: mRNA  
 A:Residues: 1-108 <MA2>  
 A:Cross-references: UNIPARC:UPI0000003F65; EMBL:X58380; NID:951343; PIDN:CAA41270.1; PIDN:R:Gardner, J.; Sautiere, P.; Cary, P.; Crai  
 C:Comment: This protein is a small, highly charged protein, and contains an important con  
 ctor in combinatorial promoters, and involved in the process of cell transformation. It i  
 C:Genetics:  
 A:Gene: Hmgi-c  
 A:Introns: 37/3; 66/3; 83/3; 94/3  
 C:Superfamily: nonhistone chromosomal protein HMG-I  
 C;Keywords: DNA binding; phosphoprotein





hypothetical protein a334R - Chlorella virus PBCV-1  
C:Species: Chlorella virus PBCV-1  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T17833  
R:Graves, M.V.; Van Etten, J.L.  
submitted to the EMBL Data Library, May 1999  
A:Reference number: Z18806  
A:Accession: T17833  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-110 <GRA>  
A:Cross-references: UNIPROT:Q84648; UNIPARC:UPI00000F752F; EMBL:U42580; NID:g4028896; PI  
A:Experimental source: specific host Chlorella strain NC64A  
C:Genetics:  
A:Note: a334R

Query Match 100.0%; Score 27; DB 2; Length 110;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
DB 52 RGRGR 56

RESULT 19  
S57448  
DNA binding protein - rice (fragment)  
C:Species: Oryza sativa (rice)  
C>Date: 10-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 09-Jul-2004  
C:Accession: S57448  
R:Meijer, A.H.; Hoge, J.H.C.  
submitted to the EMBL Data Library, June 1995  
A:Description: Three AT hook-containing proteins fom rice bind recognition sites of plan  
A:Reference number: S57448  
A:Accession: S57448  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-112 <MEI>  
A:Cross-references: UNIPROT:Q40726; UNIPARC:UPI00000A3EAB; EMBL:X88799; NID:g871497; PID

Query Match 100.0%; Score 27; DB 2; Length 112;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
DB 12 RGRGR 16

RESULT 20  
D69188  
nitrogen regulatory protein P-II MTH664 - Methanobacterium thermoautotrophicum (strain D  
C:Species: Methanobacterium thermoautotrophicum  
A:Variety: strain Delta H  
C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004  
C:Accession: D69188  
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;  
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, N.  
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
J. Bacteriol. 179, 7135-7155, 1997  
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct  
A:Reference number: A69000; MUID:98037514; PMID:9371463  
A:Accession: D69188  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-115 <MTH>  
A:Cross-references: UNIPROT:Q26760; UNIPARC:UPI000012B6AF; GB:AE0000666; NID  
A:Experimental source: strain Delta H  
C:Genetics:  
A:Gene: MTH664  
A:Start codon: GTG  
C:Superfamily: regulatory protein P-II

C:Keywords: phosphoprotein; signal transduction  
F:54/Binding site: UMP (Tyr) (covalent) #status predicted

Query Match 100.0%; Score 27; DB 2; Length 115;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
DB 37 RGRGR 41

RESULT 21  
B85036  
small nuclear riboprotein Sm-D1-like protein [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
C:Accession: B85036  
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring  
Nature 402, 769-777, 1999  
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A:Reference number: A85001; MUID:20083488; PMID:10617198  
A:Accession: B85036  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-116 <STO>  
A:Cross-references: UNIPROT:Q9SY09; UNIPARC:UPI000005SEC00; GB:NC\_001268; NID:g7269769; P  
C:Genetics:  
A:Gene: AT4G02840  
A:Map position: 4

Query Match 100.0%; Score 27; DB 2; Length 116;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
DB 101 RGRGR 105

RESULT 22  
T38440  
small nuclear ribonucleoprotein smd1 - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: T38440  
R:McDougall, R.; Wood, V.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, December 1997  
A:Reference number: Z21793  
A:Accession: T38440  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-117 <MCD>  
A:Cross-references: UNIPROT:Q42661; UNIPARC:UPI0000135A97; EMBL:AL009227; PIDN:CAA15826.1  
A:Experimental source: strain 972h-; cosmid c27D7  
C:Genetics:  
A:Gene: SPDB:SPAC27D7.07C  
A:Map position: 1  
A:Introns: 5/2; 22/1

Query Match 100.0%; Score 27; DB 2; Length 117;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
DB 96 RGRGR 100

RESULT 23  
A27668  
Sm-D ribonucleoprotein autoantigen - human  
C:Species: Homo sapiens (man)

C;Date: 19-Nov-1988 #sequence\_revision 19-Nov-1988 #text\_change 09-Jul-2004  
C;Accession: A94201; B94201; A94626; B60532; A27668  
R;Rokeach, L.A.; Haselby, J.A.; Hoch, S.O.  
Proc. Natl. Acad. Sci. U.S.A. 85, 4832-4836, 1988  
A;Title: Molecular cloning of a cDNA encoding the human Sm-D autoantigen.  
A;Reference number: A94201; MUID:88263041; PMID:3260384  
C;Accession: A94201  
A;Molecule type: mRNA  
A;Residues: 1-63, 'R', '65-78', 'IR', '81-119 <ROK>  
A;Cross-references: UNIPROT:P13641; UNIPARC:UPI000001428D0  
A;Accession: B94201  
A;Molecule type: protein  
A;Residues: 1-11 <RO2>  
A;Cross-references: UNIPARC:UPI0000179693  
R;Rokeach, L.A.  
submitted to the Protein Sequence Database, July 1988  
A;Reference number: A94626  
A;Accession: A94626  
A;Molecule type: mRNA  
A;Residues: 64, 79-80 <RO3>  
A;Cross-references: UNIPARC:UPI000011E97E  
R;Renz, M.; Heim, C.; Braeunling, O.; Czichos, A.; Wieland, C.; Seelig, H.P.  
Clin. Chem. 35, 1861-1863, 1989  
A;Title: Expression of the major human ribonucleoprotein (RNP) autoantigens in Escherichia coli.  
A;Reference number: A60532; MUID:89376817; PMID:2528429  
A;Accession: B60532  
A;Status: not compared with conceptual translation  
A;Molecule type: mRNA  
A;Residues: 1-119 <REN>  
A;Cross-references: UNIPARC:UPI0000006BEC

Query Match 100.0%; Score 27; DB 2; Length 119;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
Db 98 RGRGR 102

RESULT 24  
T40190  
probable small ribonuclear protein-sm like - fission yeast (Schizosaccharomyces pombe)  
C;Species: Schizosaccharomyces pombe  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C;Accession: T40190  
R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Lauber, J.; Hilbert, H.; Duesterhoeft, A.  
submitted to the EMBL Data Library, February 1998  
A;Reference number: Z21910  
A;Accession: T40190  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-121 <WOO>  
A;Cross-references: UNIPROT:O14352; UNIPARC:UPI000012B967; EMBL:Z97992; PIDN: CAB10801.1;  
A;Experimental source: strain 972h-; cosmid c30D10  
C;Genetics:  
A;Gene: SPDB:SPBC30D10.06  
A;Map position: 2  
A;Introns: 1/3; 65/3

Query Match 100.0%; Score 27; DB 2; Length 121;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
Db 113 RGRGR 117

RESULT 25  
D71355  
probable ribosomal protein L22 (rplV) - syphilis spirochete  
C;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)

C;Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 09-Jul-2004  
C;Accession: D71355  
R;Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin  
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDo  
rsey, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
Science 281, 375-388, 1998  
A;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.  
A;Reference number: A71250; MUID:98332770; PMID:9665876  
C;Accession: D71355  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-124 <COL>  
A;Cross-references: UNIPROT:O83224; UNIPARC:UPI0000133D88; GB:AE001202; GB:AE0000520; NID  
A;Experimental source: strain Nichols  
C;Genetics:  
A;Gene: TP0194  
C;Superfamily: Escherichia coli ribosomal protein L22

Query Match 100.0%; Score 27; DB 2; Length 124;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
Db 93 RGRGR 97

RESULT 26  
A32484  
basic fibroblast growth factor precursor, 25K - guinea pig (fragments)  
C;Species: Cavia porcellus (guinea pig)  
C;Date: 20-Oct-1989 #sequence\_revision 20-Oct-1989 #text\_change 15-Jun-1996  
C;Accession: A32484  
R;Sommer, A.; Moscatelli, D.; Rifkin, D.B.  
Biochem. Biophys. Res. Commun. 160, 1267-1274, 1989  
A;Title: An amino-terminally extended and post-translationally modified form of a 25kD ba  
A;Reference number: A32484; MUID:89273588; PMID:2730845  
A;Accession: A32484  
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra  
A;Molecule type: mRNA  
A;Residues: 1-125 <SOM>  
A;Cross-references: UNIPARC:UPI000017652E  
C;Superfamily: fibroblast growth factor

Query Match 100.0%; Score 27; DB 2; Length 125;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
Db 4 RGRGR 8

RESULT 27  
T16952  
hypothetical protein T28D9.10 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C;Accession: T16952  
R;Fulton, L.  
submitted to the EMBL Data Library, June 1995  
A;Description: The sequence of C. elegans cosmid T28D9.  
A;Reference number: Z18614  
A;Accession: T16952  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-126 <FUL>  
A;Cross-references: UNIPROT:Q10013; UNIPARC:UPI00000611AF; EMBL:U28738; NID:9861262; PID  
A;Experimental source: strain Bristol N2  
C;Genetics:  
A;Gene: CESP:T28D9.10  
A;Introns: 27/1; 62/1; 96/1

```
Query Match      100.0%; Score 27; DB 2; Length 126;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
   |||||
Db 98 RGRGR 102

RESULT 28
F97593
SOS ribosomal protein L22 [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
C:Accession: F97593
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorllo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: F97593
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-129 <KUR>
A:Cross-references: UNIPROT:Q8UE23; UNIPARC:UPI00000D1D1A; GB:AE007869; PIDN:AAK87703.1;
C:Genetics:
A:Gene: AGR_C_3548
A:Map position: circular chromosome
C:Superfamily: Escherichia coli ribosomal protein L22

Query Match      100.0%; Score 27; DB 2; Length 129;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
   |||||
Db 100 RGRGR 104

RESULT 29
AC2815
SOS ribosomal protein L22 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: AC2815
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AC2815
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-129 <KUR>
A:Cross-references: UNIPROT:Q8UE23; UNIPARC:UPI00000D1D1A; GB:AE008688; PIDN:AAL42937.1;
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: rplV
A:Map position: circular chromosome
C:Superfamily: Escherichia coli ribosomal protein L22

Query Match      100.0%; Score 27; DB 2; Length 129;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
   |||||
Db 100 RGRGR 104
```

```
RESULT 30
T46388
hypothetical protein DKFZp434A1820.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004
C:Accession: T46388
R:Ottewaelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23031
A:Accession: T46388
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-129 <AAA>
A:Cross-references: UNIPROT:Q9NT41; UNIPARC:UPI000006EFC5; EMBL:AL1137545
A:Experimental source: adult testis; clone DKFZp434A1820
C:Genetics:
A:Note: DKFZp434A1820.1

Query Match      100.0%; Score 27; DB 2; Length 129;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
   |||||
Db 57 RGRGR 61

RESULT 31
T31498
hypothetical protein Y116A8C.42 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C:Accession: T31498
R:McMurray, A.
submitted to the EMBL Data Library, October 1999
A:Reference number: Z21041
A:Accession: T31498
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-136 <WIL>
A:Cross-references: UNIPROT:O17348; UNIPARC:UPI0000135A9B; EMBL:AL117204; PIDN:CA855132.1
A:Experimental source: clone Y116A8C
C:Genetics:
A:Gene: CESP:Y116A8C.42
A:Introns: 26/1; 68/3

Query Match      100.0%; Score 27; DB 2; Length 136;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
   |||||
Db 105 RGRGR 109

RESULT 32
T43028
HMG-protein I alpha chain - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004
C:Accession: T43028; T16801
R:Kurz, T.; Schulze, E.
submitted to the EMBL Data Library, April 1998
A:Description: The high mobility group proteins of Caenorhabditis elegans.
A:Reference number: Z22282
A:Accession: T43028
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-137 <KUR>
A:Cross-references: UNIPROT:Q22204; UNIPARC:UPI0000075B7D; EMBL:AF056578; PIDN:AACT86600.1
R:Chissoe, S.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid T05A7.
```

```
A:Reference number: 218580
A:Accession: T16801
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-137 <CHI>
A:Cross-references: UNIPARC:UPI0000075B7D; EMBL:U40028; NID:G1055143; PID:G1055146; PIDN:
C:Genetics:
A:Gene: hmg-1-alpha; T05A7.4
A:Introns: 29/1

Query Match 100.0%; Score 27; DB 2; Length 137;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
DB 38 RGRGR 42

RESULT 33
insulin-like growth factor precursor - Atlantic hagfish (fragment)
A:Accession: A38612
C:Species: Myxine glutinosa (Atlantic hagfish)
C:Date: 23-Aug-1991 #sequence_revision 20-Sep-1991 #text_change 09-Jul-2004
C:Accession: A38612
R:Nagamatsu, S.; Chan, S.J.; Falkmer, S.; Steiner, D.F.
J. Biol. Chem. 266, 2397-2402, 1991
A:Title: Evolution of the insulin gene superfamily. Sequence of a preproinsulin-like gro
A:Reference number: A38612; MUID:91115860; PMID:1989990
A:Accession: A38612
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-139 <NAG>
A:Cross-references: UNIPROT:P22618; UNIPARC:UPI000012D414; GB:N57735
A:Note: the authors translated the codon TGC for residue 21 as Ser
C:Superfamily: insulin

Query Match 100.0%; Score 27; DB 2; Length 139;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
DB 133 RGRGR 137

RESULT 34
T39554
very hypothetical protein SPBC16C6.03c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T39554
R:Purnelle, B.; Goffeau, A.; Wood, V.; Lyne, M.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, February 1998
A:Reference number: Z21863
A:Accession: T39554
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-141 <PUR>
A:Cross-references: UNIPROT:O42927; UNIPARC:UPI0000179C10; EMBL:AL021767; PIDN:CAA16911.
A:Experimental source: strain 972h; cosmid c16C6
C:Genetics:
A:Gene: SPDB:SPBC16C6.03c
A:Map position: 2

Query Match 100.0%; Score 27; DB 2; Length 141;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
DB 65 RGRGR 69
```

## RESULT 35

## BVECTR

tolR protein - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C:Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 09-Jul-2004

C:Accession: C25980; A64810

R:Sun, T.P.; Webster, R.E.

J. Bacteriol. 169, 2667-2674, 1987

A:Title: Nucleotide sequence of a gene cluster involved in entry of E colicins and single

A:Reference number: A91835; MUID:87222192; PMID:3294803

A:Accession: C25980

A:Molecule type: DNA

A:Residues: 1-142 <SUN>

A:Cross-references: UNIPROT:P05829; UNIPARC:UPI0000137114; GB:M16489; NID:G148021; PIDN:R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Coi

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: A64810

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-142 <BLAT>

A:Cross-references: UNIPARC:UPI0000137114; GB:AE000177; GB:U00096; NID:G1786955; PIDN:AA

A:Experimental source: strain K-12, substrain MG1655

C:Comment: This is one of the proteins, encoded by the fli-tolAB gene cluster, that is r

C:Genetics:

A:Gene: tolR

A:Map position: 17 min

C:Superfamily: tolR protein

C:Keywords: transmembrane protein

F:18-34/Domain: transmembrane #status predicted <TMM>

Query Match 100.0%; Score 27; DB 1; Length 142;

Best Local Similarity 100.0%; Pred. No. 1.6e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5

DB 5 RGRGR 9

## RESULT 36

## AF0592

tolR protein [imported] - Salmonella enterica subsp. enterica serovar Typhi (strain CT18)

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhi

C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002

C:Accession: AF0592

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AF0592

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-142 <PAR>

A:Cross-references: UNIPARC:UPI000005A116; GB:AL513382; PIDN:CAD05208.1; PID:G16501978;

C:Genetics:

A:Gene: STY0792

C:Superfamily: tolR protein

Query Match 100.0%; Score 27; DB 2; Length 142;

Best Local Similarity 100.0%; Pred. No. 1.6e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5

DB 5 RGRGR 9

```

RESULT 37
F85576
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: F85576
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: F85576
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-142 <STO>
A:Cross-references: UNIPROT:P05829; UNIPARC:UPI0000137114; GB:AE005174; NID:g12513671; B
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: tolR
C:Superfamily: tolR protein

Query Match      100.0%; Score 27; DB 2; Length 142;
Best Local Similarity 100.0%; Pred. No. 1.6e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 5; Conservative 0;

QY      1 RGRGR 5
        |||||
Db       5 RGRGR 9

RESULT 38
E90725
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: E90725
R:Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: E90725
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-142 <HAY>
A:Cross-references: UNIPROT:P05829; UNIPARC:UPI0000137114; GB:BA000007; PIDN:BA034196.1;
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: ECs0773
C:Superfamily: tolR protein

Query Match      100.0%; Score 27; DB 2; Length 142;
Best Local Similarity 100.0%; Pred. No. 1.6e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 5; Conservative 0;

QY      1 RGRGR 5
        |||||
Db       5 RGRGR 9

RESULT 39
S22311
HMG-Y-related protein (variant B) - soybean (fragment)
C:Species: Glycine max (soybean)
C:Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C:Accession: S22311
R:Laux, T.; Seurinck, J.; Goldberg, R.B.
Nucleic Acids Res. 19, 4768, 1991
A:Title: A soybean embryo cDNA encodes a DNA binding protein with histone and HMG-protein
A:Reference number: S22310; MUID:91367678; PMID:1891368
A:Accession: S22311
```

```

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-152 <LAU>
A:Cross-references: UNIPROT:Q10370; UNIPARC:UPI000012CA2D; EMBL:X58244; NID:g18648; PIDN:
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1991
C:Superfamily: histone H1

Query Match      100.0%; Score 27; DB 2; Length 152;
Best Local Similarity 100.0%; Pred. No. 1.7e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 5; Conservative 0;

QY      1 RGRGR 5
        |||||
Db       61 RGRGR 65

RESULT 40
T04538
hypothetical protein F28J12.80 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004
C:Accession: T04538
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft,
submitted to the Protein Sequence Database, February 1998
A:Reference number: Z15377
A:Accession: T04538
A:Molecule type: DNA
A:Residues: 1-154 <BEV>
A:Cross-references: UNIPROT:O49512; UNIPARC:UPI0000A03E9; EMBL:AL021710
A:Experimental source: cultivar Columbia; BAC clone F28J12
C:Genetics:
A:Map position: 4
A:Introns: 58/3
A:Note: F28J12.80
C:Superfamily: Arabidopsis thaliana hypothetical protein T21C14.30

Query Match      100.0%; Score 27; DB 2; Length 154;
Best Local Similarity 100.0%; Pred. No. 1.7e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 5; Conservative 0;

QY      1 RGRGR 5
        |||||
Db       15 RGRGR 19

RESULT 41
G72249
ribosomal protein L22 - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C:Accession: G72249; S40193
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey,
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: G72249
A:Molecule type: DNA
A:Residues: 1-159 <ARN>
A:Cross-references: UNIPROT:P38511; UNIPARC:UPI0000133DB9; GB:AE001798; GB:AE000512; NID:
A:Experimental source: strain MSB8
R:Sanangelantoni, A.; Tiboni, O.
submitted to the EMBL Data Library, February 1993
A:Reference number: S37489
A:Accession: S40193
A:Molecule type: DNA
A:Residues: 1-81, r', 83-159 <SAN>
A:Cross-references: UNIPARC:UPI0000017049C; EMBL:221677; NID:g437921; PID:g437928
C:Genetics:
A:Gene: rplV; TM1495
C:Superfamily: Escherichia coli ribosomal protein L22
C:Keywords: protein biosynthesis; ribosome
```

```
Query Match      100.0%; Score 27; DB 2; Length 159;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 RGRGR 5
DB      116 RGRGR 120
      |||||

RESULT 42
T39586
rna binding protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T39586
R:Volckaert, G.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL data Library, August 1997
A:Reference number: Z21865
A:Accession: T39586
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-166 <VOL>
A:Cross-references: UNIPROT:O14327; UNIPARC:UPI0000068629; EMBL:Z99759; PIDN:CAB16904.1;
A:Experimental source: strain 972h-; cosmid c1659
C:Genetics:
A:Gene: SPDB:SPBC16B9.12c
A:Map position: 2
A:Introns: 12/3; 97/2; 126/3

Query Match      100.0%; Score 27; DB 2; Length 166;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 RGRGR 5
DB      139 RGRGR 143
      |||||

RESULT 43
T18691
hypothetical protein B0285.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18691
R:Sulston, J.
submitted to the EMBL data Library, June 1994
A:Reference number: Z19007
A:Accession: T18691
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-166 <WIL>
A:Cross-references: UNIPROT:P46553; UNIPARC:UPI000013B6DC; EMBL:Z34533; PIDN:CAA84296.1;
A:Experimental source: clone B0285
C:Genetics:
A:Gene: CESP:B0285.3
A:Map position: 3
A:Introns: 16/1; 114/3; 149/3

Query Match      100.0%; Score 27; DB 2; Length 166;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 RGRGR 5
DB      93 RGRGR 97
      |||||

RESULT 44
S43476
histone-like DNA-binding protein PF 1 - oat (strain Gary)
C:Species: Avena sativa (oat)
A:Variety: Gary
```

```
C:Date: 20-Oct-1994 #sequence_revision 19-Apr-1996 #text_change 09-Jul-2004
C:Accession: S43476
R:Nieto-Sotelo, J.; Ichida, A.; Quail, P.H.
Nucleic Acids Res. 22, 1115-1116, 1994
A:Title: Positive factor 1 (PF1) from oat is an HMGY- and H1 histone-like protein that b
A:Reference number: S43476; MUID:94203798; PMID:8152915
A:Accession: S43476
A:Molecule type: mRNA
A:Residues: 1-170 <NIE>
A:Cross-references: UNIPROT:Q38778; UNIPARC:UPI00000A16B6; EMBL:L24391; NID:g454278; PIDN
A:Experimental source: Gary
C:Genetics:
A:Gene: PF-1
C:Superfamily: histone H1

Query Match      100.0%; Score 27; DB 2; Length 170;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 RGRGR 5
DB      88 RGRGR 92
      |||||

RESULT 45
AC2855
ECF family sigma factor [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: AC2855
R:Wood, D.W.; Scubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
erge, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AC2855
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-171 <KUR>
A:Cross-references: UNIPROT:Q8UD59; UNIPARC:UPI00001646B3; GB:AE008688; PIDN:AAL43257.1;
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: SigD
A:Map position: circular chromosome

Query Match      100.0%; Score 27; DB 2; Length 171;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 RGRGR 5
DB      82 RGRGR 86
      |||||

RESULT 46
S22310
HMG-Y-related protein (variant A) - soybean
C:Species: Glycine max (soybean)
C:Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C:Accession: S22310
R:Laux, T.; Seurinck, J.; Goldberg, R.B.
Nucleic Acids Res. 19, 4768, 1991
A:Title: A soybean embryo cDNA encodes a DNA binding protein with histone and HMG-protein
A:Accession: S22310; MUID:91367678; PMID:1891368
A:Reference number:
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-176 <LAU>
A:Cross-references: UNIPROT:Q00423; UNIPARC:UPI000012CA2B; EMBL:X58246; NID:g18646; PIDN:
A:Note: the nucleotide sequence was submitted to the EMBL data Library, March 1991
```

C:Superfamily: histone H1  
C:Keywords: nucleue

Query Match 100.0%; Score 27; DB 2; Length 176;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
DB 85 RGRGR 89

RESULT 47  
A71011  
hypothetical protein PH1383 - Pyrococcus horikoshii  
C:Species: Pyrococcus horikoshii  
C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 31-Dec-2004  
C:Accession: A71011  
R:Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin  
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi  
DNA Res. 5, 55-76, 1998  
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a  
A:Reference number: A71000; MUID:98344137; PMID:9679194  
A:Accession: A71011  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-177 <KAW>  
A:Cross-references: UNIPROT:050091; UNIPARC:UPI0000062C79; GB:AP000006; NID:G3236133; PI  
A:Experimental source: strain OT3  
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank  
C:Genetics:  
A:Gene: PH1383  
C:Superfamily: uncharacterized conserved protein

Query Match 100.0%; Score 27; DB 2; Length 177;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
DB 163 RGRGR 167

RESULT 48  
T49691  
hypothetical protein B23L21.40 [imported] - Neurospora crassa  
C:Species: Neurospora crassa  
C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 18-Aug-2000  
C:Accession: T49691  
R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,  
submitted to the Protein Sequence Database, May 2000  
A:Reference number: Z25022  
A:Accession: T49691  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-177 <SCH>  
A:Cross-references: UNIPARC:UPI0000179D78; EMBL:AL356172; GSPDB:GN00116; NCSP:B23L21.40  
A:Experimental source: BAC clone B23L21; strain OR74A  
C:Genetics:  
A:Gene: NCSP:B23L21.40  
A:Map position: 6  
A:Introns: 37/3; 61/1; 74/3; 117/3; 142/3; 161/3  
C:Superfamily: Neurospora crassa hypothetical protein B23L21.40

Query Match 100.0%; Score 27; DB 2; Length 177;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
DB 104 RGRGR 108

RESULT 49  
T09585  
high mobility group protein HMGI/Y-2 - sword bean  
C:Species: Canavalia gladiata (sword bean)  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 09-Jul-2004  
C:Accession: T09585  
R:Yamamoto, S.; Minamikawa, T.  
Plant Mol. Biol. 33, 537-544, 1997  
A:Title: Two genes for the high mobility group protein HMG-Y are present in the genome of  
A:Reference number: Z16751; MUID:97201487; PMID:9049273  
A:Accession: T09585  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-178 <YAM>  
A:Cross-references: UNIPROT:Q42492; UNIPARC:UPI00000ACC56; EMBL:D86595  
C:Genetics:  
A:Introns: 17/3  
C:Superfamily: histone H1

Query Match 100.0%; Score 27; DB 2; Length 178;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
DB 87 RGRGR 91

RESULT 50  
T09584  
high mobility group protein HMGI/Y-1 - sword bean  
C:Species: Canavalia gladiata (sword bean)  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 09-Jul-2004  
C:Accession: T09584  
R:Yamamoto, S.; Minamikawa, T.  
Plant Mol. Biol. 33, 537-544, 1997  
A:Title: Two genes for the high mobility group protein HMG-Y are present in the genome of  
A:Reference number: Z16751; MUID:97201487; PMID:9049273  
A:Accession: T09584  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-178 <YAM>  
A:Cross-references: UNIPROT:Q42461; UNIPARC:UPI00000ACC4C; EMBL:D86594  
C:Genetics:  
A:Introns: 17/3  
C:Superfamily: histone H1

Query Match 100.0%; Score 27; DB 2; Length 178;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
DB 87 RGRGR 91

RESULT 51  
T39529  
hypothetical protein SPBC1685.13 - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: T39529  
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Hilbert, H.; Duesterhoeft, A.  
submitted to the EMBL Data Library, July 1998  
A:Reference number: Z21861  
A:Accession: T39529  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-183 <WOO>  
A:Cross-references: UNIPROT:O74333; UNIPARC:UPI000006BD2C; EMBL:AL031154; PIDN:CAA20061.1  
A:Experimental source: strain 972h-; cosmid c1685  
C:Genetics:  
A:Gene: SPDB:SPBC1685.13



A;Map position: 2

Query Match 100.0%; Score 27; DB 2; Length 183;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
|||||  
Db 162 RGRGR 166

RESULT 52

S47020  
ribosomal protein S13 - Sulfolobus acidocaldarius  
C;Species: Sulfolobus acidocaldarius  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
C;Accession: S47020  
R;Langer, D.; Hain, J.; Thuriaux, P.; Zillig, W.  
submitted to the EMBL Data Library, July 1994  
A;Description: Similarity of the transcription systems of Eukarya and Archaea.  
A;Reference number: S47020  
A;Accession: S47020  
A;Molecule type: DNA  
A;Residues: 1-184 <LAN>  
A;Cross-references: UNIPROT:P39470; UNIPARC:UPI0000134CB5; EMBL:X80194; NID:G517286; PID

C;Genetics:  
A;Gene: rps13  
C;Superfamily: ribosomal protein S13/S18  
C;Keywords: protein biosynthesis; ribosome

Query Match 100.0%; Score 27; DB 1; Length 184;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
|||||  
Db 110 RGRGR 114

RESULT 53

H97631  
hypothetical protein AGR\_C\_4121 [imported] - Agrobacterium tumefaciens (strain C58, Cere  
C;Species: Agrobacterium tumefaciens  
C;Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 09-Jul-2004  
C;Accession: H97631  
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,  
A.; Liu, P.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A;Reference number: A97359; MUID:21608551; PMID:11743194

A;Accession: H97631  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-184 <KUR>  
A;Cross-references: UNIPROT:Q8UD59; UNIPARC:UPI00000D1E3A; GB:AE007869; PIDN:AAK88009.1;  
C;Genetics:  
A;Gene: AGR\_C\_4121  
A;Map position: circular chromosome

Query Match 100.0%; Score 27; DB 2; Length 184;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
|||||  
Db 95 RGRGR 99

RESULT 54

C87500  
conserved hypothetical protein CC2024 [imported] - Caulobacter crescentus  
C;Species: Caulobacter crescentus  
C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 09-Jul-2004

C;Accession: C87500

R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.;  
B.; Laub, B.C.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A;Title: Complete Genome Sequence of Caulobacter crescentus.

A;Reference number: A87249; MUID:21173698; PMID:11259647  
A;Accession: C87500  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-186 <STO>  
A;Cross-references: UNIPROT:Q9A6R0; UNIPARC:UPI00000C75A6; GB:AE005673; NID:G13423497; P  
C;Genetics:  
A;Gene: CC2024

Query Match 100.0%; Score 27; DB 2; Length 186;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
|||||  
Db 68 RGRGR 72

RESULT 55

A48834  
basic fibroblast growth factor - chicken  
C;Species: Gallus gallus (chicken)  
C;Date: 01-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 31-Dec-2004  
C;Accession: A48834; S23636  
R;Borja, A.Z.; Meijers, C.; Zeller, R.  
Dev. Biol. 157, 110-118, 1993  
A;Title: Expression of alternatively spliced bFGF first coding exons and antisense mRNAs  
A;Reference number: A48834; MUID:93246053; PMID:7683281

A;Accession: A48834  
A;Status: preliminary  
A;Molecule type: nucleic acid  
A;Residues: 1-189 <BOR>

A;Cross-references: UNIPROT:Q07659; UNIPARC:UPI000017652B  
A;Experimental source: embryo

A;Note: sequence extracted from NCBI backbone (NCBIN:131000, NCBI:P:131001)  
R;Nitrani, E.; Gruenbaum, Y.; Shohat, H.; Ziv, T.  
Development 109, 387-393, 1990

A;Title: Fibroblast growth factor during mesoderm induction in the early chick embryo.  
A;Reference number: S23636; MUID:90382254; PMID:2401202

A;Accession: S23636  
A;Status: preliminary  
A;Molecule type: DNA

A;Residues: 95-128 <MIT>

A;Cross-references: UNIPARC:UPI0000171231; EMBL:X56804; NID:G62855; PIDN:CAA40139.1; PID

Query Match 100.0%; Score 27; DB 2; Length 189;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
|||||  
Db 4 RGRGR 8

RESULT 56

C72450  
probable DNA-3-methyladenine glycosidase APE2247 - Aeropyrum pernix (strain K1)  
C;Species: Aeropyrum pernix  
C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jul-2004  
C;Accession: C72450

R;Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah  
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, Y.; Kudoh, Y.; Yamazaki, J.; K  
DNA Res. 6, 83-101, 1999

A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr  
A;Reference number: A74450; MUID:99310339; PMID:10382966

A;Accession: C72450  
A;Status: preliminary

```

A:Molecule type: DNA
A:Residues: 1-194 <KAW>
A:Cross-references: UNIPROT:Q9Y9P1; UNIPARC:UPI000005E287; DDBJ:AP000064; NID:g5105945;
A:Experimental source: strain KI
C:Genetics:
C:Superfamily: Bacillus subtilis DNA-3-methyladenine glycosidase homolog yx1J
Query Match 100.0%; Score 27; DB 2; Length 194;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 RGRGR 5
Db 57 RGRGR 61

RESULT 57
D83175
hypothetical protein PA3765 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C:Accession: D83175
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: D83175
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-194 <STO>
A:Cross-references: UNIPROT:Q9HXN0; UNIPARC:UPI00000C5AFE; GB:AE004795; GB:AE004091; NID
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA3765

Query Match 100.0%; Score 27; DB 2; Length 194;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 RGRGR 5
Db 84 RGRGR 88

RESULT 58
S57948
HMG1/Y protein - garden pea
C:Species: Pisum sativum (garden pea)
C>Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 09-Jul-2004
C:Accession: S57948
R:Webster, C.I.; Packman, L.C.; Pwee, K.H.; Knight, J.S.; Gray, J.C.
submitted to the EMBL Data Library, July 1995
A:Description: HMG1 enhances binding of HMG1/Y to a positive regulatory region of the pe
A:Reference number: S57948
A:Accession: S57948
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-197 <WEB>
A:Cross-references: UNIPROT:Q43877; UNIPARC:UPI00000ACB6B; EMBL:X89568; NID:g899366; PID
C:Superfamily: histone H1

Query Match 100.0%; Score 27; DB 2; Length 197;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 RGRGR 5
Db 87 RGRGR 91

```

```

RESULT 59
T19797
hypothetical protein C36F7.4b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 31-Dec-2004
C:Accession: T19797
R:Lightning, J.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19178
A:Accession: T19797
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-198 <WIL>
A:Cross-references: UNIPROT:Q9NAR0; UNIPARC:UPI0000179FC0; EMBL:Z81045; PIDN:CAB54202.1;
A:Experimental source: clone C36F7
C:Genetics:
A:Gene: CESP:C36F7.4b
A:Map position: 1
A:Introns: 77/1, 115/3; 162/3

Query Match 100.0%; Score 27; DB 2; Length 198;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 RGRGR 5
Db 44 RGRGR 48

RESULT 60
T48099
hypothetical protein T20010.200 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 09-Jul-2004
C:Accession: T48099
R:Obermaier, B.; Ottenwaelder, B.; Duchemin, D.; Zeitler, K.; Mewes, H.W.; Rudd, S.; Lem
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z24484
A:Accession: T48099
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-199 <OBE>
A:Cross-references: UNIPROT:Q9LYB2; UNIPARC:UPI000009ED11; EMBL:AL163816
A:Experimental source: cultivar Columbia; BAC clone T20010
C:Genetics:
A:Map position: 3
A:Introns: 163/2
A:Note: T20010.200

Query Match 100.0%; Score 27; DB 2; Length 199;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 RGRGR 5
Db 135 RGRGR 139

RESULT 61
G81967
probable regulator of pilE expression NMA0498 [imported] - Neisseria meningitidis (strain
C:Species: Neisseria meningitidis
C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 12-Jul-2004
C:Accession: G81967
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morell
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: A81775; MUID:20222556; PMID:10761919
A:Accession: G81967
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-201 <PAR>

```

```

A;Cross-references: UNIPROT:Q9JW85; UNIPARC:UPI00000C49AB; GB:AL162753; GB:AL157959; NID
A;Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: regF; NMA0498
C:Superfamily: stringent starvation protein A

Query Match      100.0%; Score 27; DB 2; Length 201;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 RGRGR 5
Db      87 RGRGR 91

RESULT 62
B81024
stringent starvation protein A NMB1953 [imported] - Neisseria meningitidis (strain MC58)
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 12-Jul-2004
C:Accession: B81024
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: B81024
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-201 <TET>
A:Cross-references: UNIPROT:Q9JXN8; UNIPARC:UPI00000C4825; GB:AE002543; GB:AE002098; NID
A;Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1953
C:Superfamily: stringent starvation protein A

Query Match      100.0%; Score 27; DB 2; Length 201;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 RGRGR 5
Db      87 RGRGR 91

RESULT 63
D69321
conserved hypothetical protein AF0572 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: D69321
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Usterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: D69321
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-202 <KLE>
A:Cross-references: UNIPROT:O29683; UNIPARC:UPI000005706C; GB:AE001065; GB:AE000782; NID
C:Superfamily: kinase with amino acid kinase domain

Query Match      100.0%; Score 27; DB 1; Length 202;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 RGRGR 5

```

```

Db      189 RGRGR 193

RESULT 64
T24537
hypothetical protein T05F1.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T24537
R;Burton, J.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19905
A:Accession: T24537
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-203 <WIL>
A:Cross-references: UNIPROT:O18036; UNIPARC:UPI0000077C3C; EMBL:Z81586; PIDN:CA804691.1;
A;Experimental source: clone T05F1
C:Genetics:
A:Gene: CESP:T05F1.4
A:Map position: 1
A:Introns: 47/2; 96/1; 159/1; 194/3

Query Match      100.0%; Score 27; DB 2; Length 203;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 RGRGR 5
Db      17 RGRGR 21

RESULT 65
B83495
conserved hypothetical protein PA1198 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C:Accession: B83495
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathog
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: B83495
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-205 <STO>
A:Cross-references: UNIPROT:Q914E0; UNIPARC:UPI00000C52A1; GB:AE004550; GB:AE004091; NID:
A;Experimental source: strain PA01
C:Genetics:
A:Gene: PA1198
C:Superfamily: conserved hypothetical protein HI1314

Query Match      100.0%; Score 27; DB 2; Length 205;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 RGRGR 5
Db      144 RGRGR 148

RESULT 66
JC1235
transcription factor BTF3a - human
N;Contains: transcription factor BTF3b
C:Species: Homo sapiens (man)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jul-2004
C:Accession: JC1235; JC1236; S09160; S09161; S34664
R;Kanno, M.; Chalut, C.; Egly, J.M.
Gene 117, 219-228, 1992

```

A:Title: Genomic structure of the putative BTF3 transcription factor.

A:Reference number: JCI235; MUID:92347656; PMID:1386332

A:Accession: JCI235

A:Molecule type: DNA

A:Residues: 1-206 <KAN>

A:Cross-references: UNIPROT:P20290; UNIPARC:UPI0000177DB1; GB:M90352; GB:M90357

A:Note: the authors translated the codon GAG for residue 41 as Gln

A:Accession: JCI236

A:Molecule type: DNA

A:Residues: 45-206 <KA2>

A:Cross-references: UNIPARC:UPI000000067D; GB:M90352; GB:M90357

R:Zheng, X.M.; Black, D.; Chambon, P.; Egly, J.M.

Nature 344, 556-559, 1990

A:Title: Sequencing and expression of complementary DNA for the general transcription factor

A:Reference number: S09160; MUID:90206080; PMID:2320128

A:Accession: S09160

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-40, Q', 42-206 <ZH2>

A:Cross-references: UNIPARC:UPI0000049C1C; GB:X53280; NID:g29504; PIDN:CAA37375.1; PID:g

A:Note: part of this sequence was confirmed by amino acid sequencing

A:Accession: S09161

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 45-206 <ZH3>

A:Cross-references: UNIPARC:UPI000000067D; GB:X53281; NID:g29506; PIDN:CAA37376.1; PID:g

A:Note: part of this sequence was confirmed by amino acid sequencing

R:Leifer, H.; Honore, B.; Madsen, A.; Nielsen, M.S.; Anderson, A.H.; Celis, J.E.

submitted to the EMBL Data Library, July 1993

A:Description: cDNA expression and human 2D-gel data bases: towards integrating protein

A:Reference number: S34664

A:Accession: S34664

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 45-206 <LEF>

A:Cross-references: UNIPARC:UPI000000067D; EMBL:X74070; NID:g395086; PIDN:CAA52200.1; PI

C:Genetics:

A:Gene: GDB:BTF3

A:Cross-references: GDB:I35165

A:Map position: 9q13-9q13

A:Introns: 44/3; 67/3; 105/3; 173/1; 191/1

C:Superfamily: transcription factor BTF3

C:Keywords: alternative splicing; transcription factor

Query Match 100.0%; Score 27; DB 2; Length 206;

Best Local Similarity 100.0%; Pred. No. 2.2e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5

DB 13 RGRGR 17

RESULT 67

T21689

Hypothetical protein F33A8.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

A:Accession: T21689

R:Matthews, L.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19459

A:Accession: T21689

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-208 <WIL>

A:Cross-references: UNIPROT:O62213; UNIPARC:UPI000007EDF0; EMBL:Z81525; PIDN:CAB04257.1;

A:Experimental source: clone F33A8

C:Genetics:

A:Gene: CESP:F33A8.3

A:Map position: 2

A:Introns: 5/3; 174/3

C:Superfamily: Arabidopsis glycine-rich protein 2; cold shock domain homology

F:23-87/Domain: cold shock domain homology <CSD>

Query Match 100.0%; Score 27; DB 2; Length 208;

Best Local Similarity 100.0%; Pred. No. 2.2e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5

DB 116 RGRGR 120

RESULT 68

C84404

Hypothetical protein Vng2543c [imported] - Halobacterium sp. NRC-1

C:Species: Halobacterium sp. NRC-1

C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004

C:Accession: C84404

R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.

; Lethauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablor

Jung, K.H.; Alam, M.; Freitas, T.

Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000

A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li;

A:Title: Genome sequence of Halobacterium species NRC-1.

A:Reference number: A84160; MUID:20504483; PMID:11016950

A:Accession: C84404

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-210 <STO>

A:Cross-references: UNIPROT:Q9HMH2; UNIPARC:UPI0000063B73; GB:AE004437; NID:g10581934; P;

C:Genetics:

A:Gene: VNG2543C

C:Superfamily: Pyrococcus horikoshii hypothetical protein PH0010

Query Match 100.0%; Score 27; DB 2; Length 210;

Best Local Similarity 100.0%; Pred. No. 2.2e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5

DB 58 RGRGR 62

RESULT 69

B75450

Hypothetical protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004

C:Accession: B75450

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

; M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma;

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: B75450

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-210 <WHI>

A:Cross-references: UNIPROT:Q9RVL7; UNIPARC:UPI00000C1867; GB:AE001952; GB:AE000513; NID:

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR1010

A:Map position: 1

Query Match 100.0%; Score 27; DB 2; Length 210;

Best Local Similarity 100.0%; Pred. No. 2.2e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5

DB 142 RGRGR 146

## RESULT 70

T41982  
 C:Species: human herpesvirus 7 (strain JI) (fragment)  
 A:Variety: strain JI  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
 C:Accession: T41982  
 R:Nicholas, J.  
 submitted to the EMBL Data Library, December 1995  
 A:Description: Determination and analysis of the complete nucleotide sequence of human h  
 A:Reference number: Z22022  
 A:Accession: T41982  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-210 <NIC>  
 A:Cross-references: UNIPROT:Q69511; UNIPARC:UPI00000F673A; EMBL:U43400; PIDN:AAAC54742.1  
 A:Experimental source: strain JI  
 C:Genetics:  
 A:Note: U80

Query Match 100.0%; Score 27; DB 2; Length 210;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5

|||||  
 Db 161 RGRGR 165

## RESULT 71

A32398  
 basic fibroblast growth factor precursor, 22.5K form - human  
 N:Alternate names: bFGF; fibroblast growth factor 2; prostatic growth factor; prostatic  
 N:Contains: basic fibroblast growth factor, 18K form  
 C:Species: Homo sapiens (man)  
 C:Date: 31-Jul-1989 #sequence\_revision 31-Dec-1993 #text\_change 31-Dec-2004  
 C:Accession: A32398; A61537; A26642; B32878; S00297; A54316; B54316; A33624; A25824; B24  
 R:Prats, H.; Kaghad, M.; Prats, A.C.; Klagsbrun, M.; Lelias, J.M.; Liauzun, P.; Chalon,  
 Proc. Natl. Acad. Sci. U.S.A. 86, 1836-1840, 1989  
 A:Title: High molecular mass forms of basic fibroblast growth factor are initiated by al  
 A:Reference number: A32398; MUID:89184522; PMID:2338817  
 A:Accession: A32398  
 A:Molecule type: mRNA  
 A:Residues: 1-210 <PRA>  
 A:Cross-references: UNIPROT:P09038; UNIPROT:Q9UCS6; UNIPROT:Q9UCS5; UNIPROT:O00527; UNIP  
 R:Shibata, F.; Baird, A.; Florjanczyk, R.Z.  
 Growth Factors 4, 277-287, 1991

A:Title: Functional characterization of the human basic fibroblast growth factor gene pr  
 A:Reference number: A61537; MUID:92110035; PMID:1764264  
 A:Accession: A61537  
 A:Molecule type: DNA  
 A:Residues: 1-114 <SHI>  
 A:Cross-references: UNIPARC:UPI0000073DBE  
 A:Note: authors translated the codon GGA for residue 47 as Ala  
 R:Kurokawa, T.; Sasada, R.; Iwane, M.; Igarashi, K.  
 FEBS Lett. 213, 189-194, 1987  
 A:Title: Cloning and expression of cDNA encoding human basic fibroblast growth factor.  
 A:Reference number: A26642; MUID:87162468; PMID:2435575  
 A:Accession: A26642

A:Molecule type: mRNA  
 A:Residues: 56-210 <KUR>  
 A:Cross-references: UNIPARC:UPI000002COA8; GB:M27968; NID:g182562; PIDN:AAAS2448.1; PID:  
 R:Abraham, J.A.; Whang, J.L.; Tumolo, A.; Mergia, A.; Fiddes, J.C.  
 Cold Spring Harb. Symp. Quant. Biol. 51, 657-668, 1986  
 A:Title: Human basic fibroblast growth factor: nucleotide sequence, genomic organization  
 A:Reference number: A90924; MUID:87217066; PMID:3472745  
 A:Accession: B32878

A:Molecule type: mRNA  
 A:Residues: 56-210 <ABR>  
 A:Cross-references: UNIPARC:UPI000002COA8  
 A:Note: the authors translated the codon GAA for residue 108 as Gly  
 R:Abraham, J.A.; Whang, J.L.; Tumolo, A.; Mergia, A.; Friedman, J.; Gospodarowicz, D.; F  
 EMBO J. 5, 2523-2528, 1986

A:Title: Human basic fibroblast growth factor: nucleotide sequence and genomic organization;  
 A:Reference number: S00297; MUID:87053817; PMID:3780670  
 A:Accession: S00297  
 A:Status: not compared with conceptual translation  
 A:Molecule type: DNA  
 A:Residues: 1-155 <NB2>  
 A:Cross-references: UNIPARC:UPI000017652F  
 A:Note: the authors translated the codon GAA for residue 108 as Gly  
 R:Shimoyama, Y.; Gotoh, M.; Ino, Y.; Sakamoto, M.; Kato, K.; Hirohashi, S.  
 Jpn. J. Cancer Res. 82, 1263-1270, 1991  
 A:Title: Characterization of high-molecular-mass forms of basic fibroblast growth factor  
 A:Accession: B54316  
 A:Molecule type: protein  
 A:Residues: 'XX', '86-88', 'X', '90-91', 'X', '93-95' <SH3>  
 A:Cross-references: UNIPARC:UPI000006E614  
 A:Experimental source: C-Li21 hepatocellular carcinoma cell line  
 A:Note: sequence extracted from NCBI backbone (NCBI:71595)  
 A:Accession: B54316  
 A:Molecule type: protein  
 A:Residues: 'XX', '19', 'X', '21-29' <SH2>  
 A:Cross-references: UNIPARC:UPI00000726DF  
 A:Note: sequence extracted from NCBI backbone (NCBI:71594)  
 R:Feige, J.J.; Bradley, J.D.; Fryburg, K.; Farris, J.; Cousins, L.C.; Barr, P.J.; Baird,  
 J. Cell Biol. 109, 3105-3114, 1989  
 A:Title: Differential effects of heparin, fibronectin, and laminin on the phosphorylation  
 A:Reference number: A33624; MUID:90078343; PMID:2592418  
 A:Accession: A33624  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 57-210 <FEI>  
 A:Cross-references: UNIPARC:UPI0000157701  
 R:Story, M.T.; Esch, F.; Shimasaki, S.; Sasse, J.; Jacobs, S.C.; Lawson, R.K.  
 Biochem. Biophys. Res. Commun. 142, 702-709, 1987  
 A:Title: Amino-terminal sequence of a large form of basic fibroblast growth factor isolat  
 A:Reference number: A25824; MUID:87156686; PMID:2435284  
 A:Accession: A25824  
 A:Molecule type: protein  
 A:Residues: 57-77 <STO>  
 A:Cross-references: UNIPARC:UPI0000176530  
 A:Experimental source: prostate  
 R:Gimenez-Gallego, G.; Conn, G.; Hatcher, V.B.; Thomas, K.A.  
 Biochem. Biophys. Res. Commun. 135, 541-548, 1986  
 A:Title: Human brain-derived acidic and basic fibroblast growth factors: amino terminal  
 A:Reference number: A90122; MUID:86186784; PMID:3364259  
 A:Accession: B24243  
 A:Molecule type: protein  
 A:Residues: 65-102, 'X', '104-105' <GIM>  
 A:Cross-references: UNIPARC:UPI0000176531  
 A:Experimental source: brain  
 R:Gautschi, P.; Frater-Schroder, M.; Bohlen, P.  
 FEBS Lett. 204, 203-207, 1986  
 A:Title: Partial molecular characterization of endothelial cell mitogens from human brain  
 A:Reference number: A91364; MUID:86275260; PMID:3732516  
 A:Accession: B24301  
 A:Molecule type: protein  
 A:Residues: 65-88, 'X', '90-98', 'X', '100' <GAU>  
 A:Cross-references: UNIPARC:UPI0000176532  
 R:Sommer, A.; Brewer, M.T.; Thompson, R.C.; Moscatelli, D.; Presta, M.; Rifkin, D.B.  
 Biochem. Biophys. Res. Commun. 144, 543-550, 1987  
 A:Title: A form of human basic fibroblast growth factor with an extended amino terminus.  
 A:Reference number: S42242; MUID:87213238; PMID:3579930  
 A:Accession: S42242  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 54-210 <SOM>  
 A:Cross-references: UNIPARC:UPI000014258B; EMBL:M17599; NID:g183086; PIDN:AAAS2534.1; PI  
 R:Pantoliano, M.W.; Horlick, R.A.; Springer, B.A.; Van Dyk, D.E.; Tobery, T.; Wetmore, D.  
 Biochemistry 33, 10229-10248, 1994  
 A:Title: Multivalent ligand-receptor binding interactions in the fibroblast growth factor;  
 A:Reference number: A55784; MUID:94347757; PMID:7520751  
 A:Accession: B55784

A:Molecule type: protein  
A:Residues: 54-71 <PAN>  
A:Cross-references: UNIPARC:UPI0000176533  
R:Watson, R.; Anthony, F.; Pickett, M.; Lambden, P.; Masson, G.M.; Thomas, B.J.  
Biochem. Biophys. Res. Commun. 187, 1227-1231, 1992  
A:Title: Reverse transcription with nested polymerase chain reaction shows expression of  
tients.  
A:Reference number: I52267; MUID:93038590; PMID:1417798  
A:Accession: I52267  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 95-182 <RES>  
A:Cross-references: UNIPARC:UPI000016B39E; GB:S47380; NID:g256535; PIDN:AADI3853.1; PID:  
A:Experimental source: granulosa cells  
R:Patry, V.; Bugler, B.; Analric, F.; Prome, J.C.; Prats, H.  
FBS Lett. 349, 23-28, 1994  
A:Title: Purification and characterization of the 210-amino acid recombinant basic fibro  
A:Reference number: S46253; MUID:94320639; PMID:8045296  
A:Molecule type: Protein  
A:Residues: 39-53; 65-88 <PAT>  
A:Cross-references: UNIPARC:UPI000002G528; UNIPARC:UPI0000176534  
A:Note: recombinant gene expressed in Escherichia coli  
C:Genetics:  
A:Gene: GDB:FGP2; FGFB  
A:Cross-references: GDB:119910; OMIM:134920  
A:Map position: 4q25-4q27  
A:Start codon: CTG  
C:Keywords: alternative initiators; angiogenesis; growth factor; heparin binding; mitoge  
F:1-210/Product: basic fibroblast growth factor, 22.5K form #status predicted <MA2>  
F:65-210/Product: basic fibroblast growth factor, 18K form #status predicted <MAT>  
F:82-86/Region: heparin binding #status predicted  
F:171-174/Region: heparin binding #status predicted

Query Match 100.0%; Score 27; DB 2; Length 210;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
| | | | |  
DB 4 RGRGR 8

RESULT 72  
C96539  
hypotheoretical protein F14I3.10 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: C96539  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
ansen, N.F.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: C96539  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-211 <STO>  
A:Cross-references: UNIPROT:Q9SX47; UNIPARC:UPI00000A0953; GB:AE005173; NID:g5734788; PI  
C:Genetics:  
A:Gene: F14I3.10  
A:Map position: 1

Query Match 100.0%; Score 27; DB 2; Length 211;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
| | | | |  
DB 4 RGRGR 8

RESULT 75  
TI9793  
hypotheoretical protein C36F7.4a - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 31-Dec-2004  
C:Accession: TI9793  
R:Lightning, J.  
submitted to the EMBL Data Library, October 1996

Query Match 100.0%; Score 27; DB 2; Length 215;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
| | | | |  
DB 205 RGRGR 209

RESULT 74  
T49743  
probable rrm-type rna binding protein [imported] - Neurospora crassa  
N:Alternate names: protein B24B19.150  
C:Species: Neurospora crassa  
C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
C:Accession: T49743  
R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,  
submitted to the Protein Sequence Database, May 2000  
A:Reference number: Z25022  
A:Accession: T49743  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-215 <SCH>  
A:Cross-references: UNIPARC:UPI000017B4D3; EMBL:AL356192; GSPDB:GN00116; NCSP:B24B19.150  
A:Experimental source: BAC clone B24B19; strain OR74A  
C:Genetics:  
A:Gene: NCSP:B24B19.150  
A:Map position: 6  
A:Introns: 42/3; 128/2

Query Match 100.0%; Score 27; DB 2; Length 213;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
| | | | |  
DB 99 RGRGR 103

RESULT 73  
T03931  
DNA binding protein PF1 - rice  
C:Species: Oryza sativa (rice)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 09-Jul-2004  
C:Accession: T03931  
R:Nieto-Sotelo, J.; Ichida, A.; Quail, P.H.  
Plant Cell 6, 287-301, 1994  
A:Title: PF1: An A-T hook containing DNA binding protein from rice that interacts with a  
A:Reference number: Z15142; MUID:94198599; PMID:8148649  
A:Accession: T03931  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-213 <NIE>  
A:Cross-references: UNIPROT:Q43600; UNIPARC:UPI00000AA86D; EMBL:L24390; NID:g453691; PID:  
A:Experimental source: cv. Nipponbare, shoot  
C:Genetics:  
A:Gene: PF1  
C:Superfamily: histone H1  
C:Keywords: DNA binding

Query Match 100.0%; Score 27; DB 2; Length 213;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
| | | | |  
DB 99 RGRGR 103

A;Reference number: Z19178  
A;Accession: T19793  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-223 <WIL>  
A;Cross-references: UNIPROT:Q93350; UNIPARC:UPI0000179FBF; EMBL:Z81045; PIDN:CAB02817.1;  
A;Experimental source: clone C36F7  
C;Genetics:  
A;Gene: CESP:C36F7.4a  
A;Map position: 1  
A;Introns: 71/2; 102/1; 140/3; 187/3

Query Match 100.0%; Score 27; DB 2; Length 223;  
Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
| | | | |  
Db 44 RGRGR 48

Search completed: December 2, 2005, 10:07:49  
Job time : 28 secs

***This Page Blank (uspto)***



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2005, 09:42:09 ; Search time 147 Seconds  
(without alignments)  
23.998 Million cell updates/sec

Title: SEQ-RGRGR  
Perfect score: 27  
Sequence: 1 rgrgr 5

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 1000 summaries

Database : Uniprot\_05.80.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	100.0	29	2	Q700R5 sus scrofa
2	27	100.0	30	2	Q5BWT8 schistosoma
3	27	100.0	35	2	Q7TPH0 mus musculus
4	27	100.0	36	2	P78343 homo sapien
5	27	100.0	37	2	Q9J154 rat
6	27	100.0	45	2	Q56X91 arabisopsis
7	27	100.0	51	2	Q6L736 9ACTO
8	27	100.0	55	2	Q9S2H2 STRCO
9	27	100.0	56	2	Q6YYD0 ORYSA
10	27	100.0	58	2	Q69128 9GAMA
11	27	100.0	59	2	Q6ZFA3 ORYSA
12	27	100.0	59	2	Q64001 9MURI
13	27	100.0	59	2	Q6QX44 9VIRU
14	27	100.0	61	2	Q80VM0 mouse
15	27	100.0	62	2	Q7QW88 GIALA
16	27	100.0	65	2	Q6Z6R0 ORYSA
17	27	100.0	65	2	Q913P6 9BACU
18	27	100.0	67	2	Q7NZG4 CHRVO
19	27	100.0	68	1	HSP1_TACAC
20	27	100.0	68	2	Q8WRX4 DROME
21	27	100.0	68	2	Q5Z559 ORYSA
22	27	100.0	68	2	Q6ERQ6 ORYSA
23	27	100.0	68	2	Q6Z0B8 ORYSA
24	27	100.0	74	2	Q61K40 DROME
25	27	100.0	76	2	Q5VQB4 ORYSA
26	27	100.0	76	2	Q6K8Z0 ORYSA
27	27	100.0	77	2	Q4SAZ2 TETNG
28	27	100.0	79	2	Q7F7E9 ORYSA
29	27	100.0	79	2	Q8V7D0 9VIRU
30	27	100.0	79	2	Q8V7E5 9VIRU
31	27	100.0	80	2	Q8LQT9 ORYSA

32	27	100.0	82	2	Q29194_PIG	Q29194 sus scrofa
33	27	100.0	82	2	Q69144_9GAMA	Q69144 human herpe
34	27	100.0	83	2	Q5W754 ORYSA	Q5W754 oryza sativ
35	27	100.0	83	2	Q8GVG2 ORYSA	Q8GVG2 oryza sativ
36	27	100.0	83	2	Q67RR9 SYMTH	Q67RR9 symbiobacte
37	27	100.0	84	2	Q6H6S6 ORYSA	Q6H6S6 oryza sativ
38	27	100.0	84	2	Q6KLR2 ORYSA	Q6KLR2 oryza sativ
39	27	100.0	85	2	Q6Z2C1 ORYSA	Q6Z2C1 oryza sativ
40	27	100.0	85	2	Q6ZLF3 ORYSA	Q6ZLF3 oryza sativ
41	27	100.0	85	2	Q8H360 ORYSA	Q8H360 oryza sativ
42	27	100.0	86	2	Q6NIQ5 CORDI	Q6NIQ5 corynebacte
43	27	100.0	86	2	Q79VI0 CORGL	Q79VI0 corynebacte
44	27	100.0	86	2	Q8FRH8 COREF	Q8FRH8 corynebacte
45	27	100.0	87	2	Q6ERR9 ORYSA	Q6ERR9 oryza sativ
46	27	100.0	87	2	Q6K9E3 ORYSA	Q6K9E3 oryza sativ
47	27	100.0	87	2	Q6ZAG7 ORYSA	Q6ZAG7 oryza sativ
48	27	100.0	88	2	Q6ER89 ORYSA	Q6ER89 oryza sativ
49	27	100.0	89	2	Q6YYW8 ORYSA	Q6YYW8 oryza sativ
50	27	100.0	89	2	Q8LH02 ORYSA	Q8LH02 oryza sativ
51	27	100.0	89	2	Q8FQ78 COREF	Q8FQ78 corynebacte
52	27	100.0	90	2	Q84ZB3 ORYSA	Q84ZB3 oryza sativ
53	27	100.0	94	2	Q7S150 ORYSA	Q7S150 oryza sativ
54	27	100.0	94	2	Q6N9G2 RHOPA	Q6N9G2 rhodopseude
55	27	100.0	94	2	Q7NPO9 CHRVO	Q7NPO9 chromobacte
56	27	100.0	95	2	Q9X115 THEME	Q9X115 thermotoga
57	27	100.0	96	2	Q5TEU8 HUMAN	Q5TEU8 homo sapien
58	27	100.0	96	2	Q6QUF0 CANFA	Q6QUF0 canis famil
59	27	100.0	96	2	Q6QUF9 CANFA	Q6QUF9 canis famil
60	27	100.0	96	2	Q8GRR5 ORYSA	Q8GRR5 oryza sativ
61	27	100.0	96	2	Q8K1F5 RAT	Q8K1F5 rattus norv
62	27	100.0	96	2	Q6W8X3_CHICK	Q6W8X3 gallus gall
63	27	100.0	96	2	Q6UF82 9HIV1	Q6UF82 human immun
64	27	100.0	96	2	Q6UFA0 9HIV1	Q6UFA0 human immun
65	27	100.0	97	2	Q6ZBK7 ORYSA	Q6ZBK7 oryza sativ
66	27	100.0	97	2	Q6N2S4 RHOPA	Q6N2S4 rhodopseude
67	27	100.0	97	2	Q91148 9HIV1	Q91148 human immun
68	27	100.0	97	2	Q77375 9HIV1	Q77375 human immun
69	27	100.0	97	2	Q99D95 9HIV1	Q99D95 human immun
70	27	100.0	97	2	Q9WQ16 9HIV1	Q9WQ16 human immun
71	27	100.0	98	2	Q5Z2E5 NOCFA	Q5Z2E5 nocardia fa
72	27	100.0	98	2	Q9FBW0 STRCO	Q9FBW0 streptomyce
73	27	100.0	98	2	Q6JNA8 9HIV1	Q6JNA8 human immun
74	27	100.0	98	2	Q6UFK9 9HIV1	Q6UFK9 human immun
75	27	100.0	99	2	Q8U540 AGRT5	Q8U540 agrobacteri
76	27	100.0	100	2	Q59GS1 HUMAN	Q59GS1 homo sapien
77	27	100.0	100	2	Q5VNN8 ORYSA	Q5VNN8 oryza sativ
78	27	100.0	100	2	Q68896 9BETA	Q68896 human herpe
79	27	100.0	100	2	Q91132 9HIV1	Q91132 human immun
80	27	100.0	100	2	Q91136 9HIV1	Q91136 human immun
81	27	100.0	100	2	Q91140 9HIV1	Q91140 human immun
82	27	100.0	100	2	Q91144 9HIV1	Q91144 human immun
83	27	100.0	100	2	Q91152 9HIV1	Q91152 human immun
84	27	100.0	100	2	Q91162 9HIV1	Q91162 human immun
85	27	100.0	100	2	Q91166 9HIV1	Q91166 human immun
86	27	100.0	100	2	Q91170 9HIV1	Q91170 human immun
87	27	100.0	100	2	Q91180 9HIV1	Q91180 human immun
88	27	100.0	100	2	Q5VGP6 9HIV1	Q5VGP6 human immun
89	27	100.0	100	2	Q66Q74 9HIV1	Q66Q74 human immun
90	27	100.0	100	2	Q6Y8S0 9HIV1	Q6Y8S0 human immun
91	27	100.0	100	2	Q6Y8S9 9HIV1	Q6Y8S9 human immun
92	27	100.0	100	2	Q6Y8T8 9HIV1	Q6Y8T8 human immun
93	27	100.0	100	2	Q6Y8U6 9HIV1	Q6Y8U6 human immun
94	27	100.0	100	2	Q6Y8V5 9HIV1	Q6Y8V5 human immun
95	27	100.0	100	2	Q6Y8W4 9HIV1	Q6Y8W4 human immun
96	27	100.0	100	2	Q6Y8X3 9HIV1	Q6Y8X3 human immun
97	27	100.0	100	2	Q6Y8Y2 9HIV1	Q6Y8Y2 human immun
98	27	100.0	100	2	Q6Y8Z1 9HIV1	Q6Y8Z1 human immun
99	27	100.0	100	2	Q6Y900 9HIV1	Q6Y900 human immun
100	27	100.0	100	2	Q6Y909 9HIV1	Q6Y909 human immun
101	27	100.0	100	2	Q6Y918 9HIV1	Q6Y918 human immun
102	27	100.0	100	2	Q6Y927 9HIV1	Q6Y927 human immun
103	27	100.0	100	2	Q6Y945 9HIV1	Q6Y945 human immun
104	27	100.0	100	2	Q79668 9HIV1	Q79668 human immun

105	27	100.0	100	2	Q900Y9_9HIV1	Q900Y9 human immun	178	27	100.0	113	2	Q4WHK1_ASPFU	Q4whk1 aspergillus
106	27	100.0	100	2	Q900Z8_9HIV1	Q900Z8 human immun	179	27	100.0	113	2	Q9UJ34_HUMAN	Q9uj34 homo sapien
107	27	100.0	100	2	Q99DA0_9HIV1	Q99da0 human immun	180	27	100.0	113	2	Q6H587_ORYSA	Q6h587 oryza sativ
108	27	100.0	100	2	Q99DA5_9HIV1	Q99da5 human immun	181	27	100.0	113	2	Q08106_9GAMR	Q08106 spleen focu
109	27	100.0	100	2	Q99DB0_9HIV1	Q99db0 human immun	182	27	100.0	114	2	O00527_HUMAN	O00527 homo sapien
110	27	100.0	100	2	Q99DB5_9HIV1	Q99db5 human immun	183	27	100.0	114	2	Q16443_HUMAN	Q16443 homo sapien
111	27	100.0	100	2	Q99DC0_9HIV1	Q99dc0 human immun	184	27	100.0	114	2	Q7XQN4_ORYSA	Q7xqn4 oryza sativ
112	27	100.0	100	2	Q99DC5_9HIV1	Q99dc5 human immun	185	27	100.0	114	2	Q6YY16_ORYSA	Q6yy16 oryza sativ
113	27	100.0	100	2	Q99DC7_9HIV1	Q99dc7 human immun	186	27	100.0	114	2	Q9SSE1_ARATH	Q9sef1 arabidopsis
114	27	100.0	100	2	Q9YNC7_9HIV1	Q9ync7 human immun	187	27	100.0	115	1	GLNB2_METH	O26760 methanobact
115	27	100.0	100	2	Q9YNC8_9HIV1	Q9ync8 human immun	188	27	100.0	115	1	RK20_NYMAL	Q6ew30 nymphaea al
116	27	100.0	100	2	Q9YNC9_9HIV1	Q9ync9 human immun	189	27	100.0	115	2	Q4FGC6_9MAGN	Q4fgc6 nuphar adve
117	27	100.0	101	2	Q61H69_DROME	Q61h69 drosophila	190	27	100.0	115	2	Q7ZEM8_9HIV1	Q7zem8 human immun
118	27	100.0	101	2	Q5JM99_ORYSA	Q5jm99 oryza sativ	191	27	100.0	116	2	Q9SY09_ARATH	Q9sy09 arabidopsis
119	27	100.0	101	2	O6NV11_BRARE	O6nv11 brachydanio	192	27	100.0	117	1	SMD1_SCHPO	O42661 schizosacch
120	27	100.0	102	2	Q9BE83_MACFA	Q9be83 macaca fasc	193	27	100.0	117	2	Q6CDE0_YARLI	Q6cde0 yarrowia li
121	27	100.0	102	2	Q5ZDE0_ORYSA	Q5zde0 oryza sativ	194	27	100.0	117	2	Q4S881_CONBU	Q4s881 coxiella bu
122	27	100.0	102	2	Q67946_ACEBU	Q67946 acetobacter	195	27	100.0	118	2	Q5D9D2_SCHJA	Q5d9d2 schistosoma
123	27	100.0	103	2	Q4YL00_PLABE	Q4yl00 plasmodium	196	27	100.0	118	2	Q6YU92_ORYSA	Q6yu92 oryza sativ
124	27	100.0	103	2	Q6ZLC0_ORYSA	Q6zlc0 oryza sativ	197	27	100.0	118	2	Q949R3_ARATH	Q949r3 arabidopsis
125	27	100.0	103	2	O4TGB3_TETNG	O4tgb3 tetraodon n	198	27	100.0	119	1	SMD1_HUMAN	P62314 homo sapien
126	27	100.0	105	2	Q6CG15_YARLI	Q6cgs15 yarrowia li	199	27	100.0	119	1	SMD1_MOUSE	P62315 mus musculu
127	27	100.0	105	2	Q6K2T3_ORYSA	Q6k2t3 oryza sativ	200	27	100.0	119	2	Q7Z5A3_HUMAN	Q7z5a3 homo sapien
128	27	100.0	105	2	Q8LL99_AEGTA	Q8ll99 aegilops ta	201	27	100.0	119	2	Q4R5F6_MACFA	Q4r5f6 macaca fasc
129	27	100.0	105	2	Q641H4_XENLA	Q641h4 xenopus lae	202	27	100.0	119	2	Q5YES0_CHLS6	Q5yes0 chlorarachn
130	27	100.0	106	1	HMGIY_CANPA	Q6urc2 canis famil	203	27	100.0	119	2	O8JHH1_BRARE	O8jhh1 brachydanio
131	27	100.0	106	1	HMGIY_CRIGR	Q9qxp3 cricetus	204	27	100.0	119	2	Q6GR62_XENLA	Q6gr62 xenopus lae
132	27	100.0	106	1	HMGIY_HUMAN	P17096 homo sapien	205	27	100.0	119	2	Q6GLH6_XENTR	Q6glh6 xenopus tro
133	27	100.0	106	1	HMGIY_MOUSE	P17095 mus musculu	206	27	100.0	120	2	Q5YET5_CHLS6	Q5yet5 chlorarachn
134	27	100.0	106	1	Q5N813_ORYSA	Q5n813 oryza sativ	207	27	100.0	121	1	LSM4_SCHPO	O14352 schizosacch
135	27	100.0	107	2	Q61PL9_HUMAN	Q61pl9 homo sapien	208	27	100.0	121	2	Q5BBH5_EMENI	Q5bbh5 aspergillus
136	27	100.0	107	2	O5T6U9_HUMAN	Q5t6u9 homo sapien	209	27	100.0	121	2	Q52FX7_MAGGR	Q52fx7 magnaporthe
137	27	100.0	107	2	Q6FGE0_HUMAN	Q6fge0 homo sapien	210	27	100.0	121	2	Q5ZGV0_ORYSA	Q5zgv0 oryza sativ
138	27	100.0	107	2	Q9LI96_ARATH	Q9li96 arabidopsis	211	27	100.0	121	2	Q7EZ39_ORYSA	Q7ez39 oryza sativ
139	27	100.0	107	2	Q7X2E1_AERSA	Q7x2e1 aeromonas s	212	27	100.0	122	2	Q6ATQ4_ORYSA	Q6atq4 oryza sativ
140	27	100.0	107	2	Q8QMP4_AERSA	Q8qmp4 aeromonas s	213	27	100.0	122	2	Q6YZ81_ORYSA	Q6yz81 oryza sativ
141	27	100.0	107	2	Q4NRH6_9DELT	Q4nrh6 anaeromyxob	214	27	100.0	122	2	Q6Z7E4_ORYSA	Q6z7e4 oryza sativ
142	27	100.0	107	2	O8R791_RAT	O8r791 rattus norv	215	27	100.0	122	2	Q6EOI8_ORYSA	Q6eoi8 oryza sativ
143	27	100.0	107	2	O8K585_RAT	O8k585 rattus norv	216	27	100.0	122	2	Q607V3_METCA	Q607v3 methylococc
144	27	100.0	107	2	Q566K0_MOUSE	Q566k0 mus musculu	217	27	100.0	122	2	Q9CVM0_MOUSE	Q9cvw0 mus musculu
145	27	100.0	108	1	HMGIY_MOUSE	P52927 mus musculu	218	27	100.0	123	2	Q7SE68_NEUCR	Q7se68 neurospora
146	27	100.0	108	2	Q5QM77_ORYSA	Q5qm77 oryza sativ	219	27	100.0	123	2	Q86TQ6_HUMAN	Q86tq6 homo sapien
147	27	100.0	108	2	Q7EY64_ORYSA	Q7ey64 oryza sativ	220	27	100.0	123	2	Q7RST3_PLAYO	Q7rst3 plasmodium
148	27	100.0	108	2	O8W332_ORYSA	Q8w332 oryza sativ	221	27	100.0	123	2	Q4XPZ0_PLACH	Q4xpz0 plasmodium
149	27	100.0	108	2	O57LU6_SALCH	O57lu6 salmonella	222	27	100.0	123	2	Q4ZLQ8_PLABE	Q4zlq8 plasmodium
150	27	100.0	108	2	Q8VKG7_MYCTU	Q8vkq7 mycobacteri	223	27	100.0	123	2	Q6LHX1_PROPR	Q6lhx1 photobacter
151	27	100.0	108	2	Q5PNE9_SALPA	Q5pne9 salmonella	224	27	100.0	123	2	Q5ZLF9_CHICK	Q5zlf9 gallus gall
152	27	100.0	108	2	Q7CQ29_SALTY	Q7cq29 salmonella	225	27	100.0	124	1	RL22_TREPA	O32224 treponema p
153	27	100.0	108	2	O8XFD1_SALTI	O8xfdl1 salmonella	226	27	100.0	124	1	SMD1_DROME	Q9vu02 drosophila
154	27	100.0	109	1	HMGIY_HUMAN	P52926 homo sapien	227	27	100.0	124	2	Q9BXE1_HUMAN	Q9bxe1 homo sapien
155	27	100.0	109	2	O6CKA2_KLUJA	Q6cka2 kluyveromyc	228	27	100.0	124	2	Q9BXE2_HUMAN	Q9bxe2 homo sapien
156	27	100.0	109	2	Q76M67_CAPHI	Q76m67 capria hircu	229	27	100.0	124	2	Q9BXE7_HUMAN	Q9bxe7 homo sapien
157	27	100.0	109	2	Q5VMG3_ORYSA	Q5vmg3 oryza sativ	230	27	100.0	124	2	Q9BZ55_HUMAN	Q9bz55 homo sapien
158	27	100.0	109	2	Q5VR66_ORYSA	Q5vr66 oryza sativ	231	27	100.0	124	2	Q60CX3_SOLTU	Q60cx3 solanum tub
159	27	100.0	109	2	Q651Y9_ORYSA	Q651y9 oryza sativ	232	27	100.0	124	2	Q5U448_MOUSE	Q5u448 mus musculu
160	27	100.0	109	2	Q7EZ60_ORYSA	Q7ez60 oryza sativ	233	27	100.0	125	2	Q5ZF39_MAGGR	Q5zf39 magnaporthe
161	27	100.0	109	2	Q657M8_ORYSA	Q657m8 oryza sativ	234	27	100.0	125	2	Q4UIE3_THEAN	Q4uie3 thaileria a
162	27	100.0	109	2	Q73816_CHICK	Q73816 gallus gall	235	27	100.0	125	2	Q4N771_THEPA	Q4n771 theileria p
163	27	100.0	110	1	Y1266_PYRAB	Y1266 pyrococcus	236	27	100.0	125	2	Q6ZLN3_ORYSA	Q6zln3 oryza sativ
164	27	100.0	110	1	Y803_PYRHO	O58533 pyrococcus	237	27	100.0	125	2	Q4NWG5_9DELT	Q4nwg5 anaeromyxob
165	27	100.0	110	2	O57Z54_HUMAN	Q57z54 homo sapien	238	27	100.0	126	1	SMD1_CAEEL	Q10013 caenorhabdi
166	27	100.0	110	2	Q6H7Q2_ORYSA	Q6h7q2 oryza sativ	239	27	100.0	126	1	SMD3_HUMAN	P62318 homo sapien
167	27	100.0	110	2	Q75GMW_ORYSA	Q75gmw oryza sativ	240	27	100.0	126	1	SMD3_MOUSE	P62320 mus musculu
168	27	100.0	110	2	Q84648_CHVP1	Q84648 paramemium	241	27	100.0	126	1	SMD3_XENLA	P62323 xenopus lae
169	27	100.0	110	2	Q4RX93_TETNG	Q4rx93 tetraodon n	242	27	100.0	126	2	Q4IR00_GIBZE	Q4ir00 gibberella
170	27	100.0	111	1	IFIA_AERPE	P57676 aeropyrum p	243	27	100.0	126	2	Q7PT70_ANOGA	Q7pt70 anopheles g
171	27	100.0	111	2	Q67V87_ORYSA	Q67v87 oryza sativ	244	27	100.0	126	2	Q91VM2_MOUSE	Q91vm2 mus musculu
172	27	100.0	111	2	O7XU56_ORYSA	Q7xu56 oryza sativ	245	27	100.0	126	2	O5ZL58_CHICK	O5zl58 gallus gall
173	27	100.0	112	2	O5XQB9_9ARCH	Q5xqb9 uncultured	246	27	100.0	126	2	Q8AVZ5_XENLA	Q8avz5 xenopus lae
174	27	100.0	112	2	O54YV9_DICDI	Q54yv9 dictyosteli	247	27	100.0	127	1	RL22_RHOA	Q6n4t9 rhodospheo
175	27	100.0	112	2	Q4YL17_PLABE	Q4yl17 plasmodium	248	27	100.0	127	2	Q5DBH5_SCHJA	Q5dbh5 schistosoma
176	27	100.0	112	2	Q40726_ORYSA	Q40726 oryza sativ	249	27	100.0	127	2	Q5CI33_CRYHO	Q5ci33 cryptospori
177	27	100.0	113	1	Y562_PYRKO	Q5jfe37 pyrococcus	250	27	100.0	127	2	Q7XHN2_ORYSA	Q7xhn2 oryza sativ

251	27	100.0	127	2	Q61056_BRARE	Q61q56 brachydanio	324	27	100.0	143	2	Q81IT3_PLAF7	Q81it3 plasmodium
252	27	100.0	127	2	Q4RZU5_TETNG	Q4ru5 tetraodon n	325	27	100.0	143	2	Q5N7M8_ORYSA	Q5nm8 oryza sativ
253	27	100.0	127	2	Q4T029_TETNG	Q4t029 tetraodon n	326	27	100.0	143	2	Q65711_ORYSA	Q65711 oryza sativ
254	27	100.0	128	2	Q5VP12_ORYSA	Q5vp12 oryza sativ	327	27	100.0	144	2	Q75M94_HUMAN	Q75m94 homo sapien
255	27	100.0	128	2	Q5PBW8_ANAMM	Q5pbw8 anaplasma m	328	27	100.0	144	2	Q6YY76_ORYSA	Q6yy76 oryza sativ
256	27	100.0	128	2	Q825S8_STRAW	Q825s8 streptomyce	329	27	100.0	144	2	Q5XG50_XENLA	Q5xg50 xenopus lae
257	27	100.0	128	2	Q7ZVB5_BRARE	Q7zvb5 brachydanio	330	27	100.0	145	2	Q529Y9_ORYSA	Q529y9 oryza sativ
258	27	100.0	129	1	RL22_AGR5	Q8ue23 agrobacteri	331	27	100.0	145	2	Q89544_9BETA	Q89544 human herpe
259	27	100.0	129	1	RL22_BRAJU	Q8g980 brachyrihizob	332	27	100.0	146	2	Q4P617_USTMA	Q4p617 ustilago ma
260	27	100.0	129	1	RL22_BRAJU	Q8g980 bruceella eu	333	27	100.0	146	2	Q53J75_ORYSA	Q53j75 oryza sativ
261	27	100.0	129	1	RL22_RHIL0	Q98m52 rhizobium l	334	27	100.0	146	2	Q98IM9_RHIL0	Q98im9 rhizobium l
262	27	100.0	129	1	RL22_RHIL0	Q92g95 rhizobium m	335	27	100.0	146	2	Q8VDA6_RAT	Q8vda6 rattus norv
263	27	100.0	129	2	Q9NT41_HUMAN	Q9nt41 homo sapien	336	27	100.0	147	1	LSM4_ORYSA	Q9lqe6 oryza sativ
264	27	100.0	129	2	Q8NBFO_HUMAN	Q8nbfo homo sapien	337	27	100.0	147	1	Q6VB70_HHVIR	Q6vb70 human herpe
265	27	100.0	129	2	Q69IT5_ORYSA	Q69it5 oryza sativ	338	27	100.0	148	1	LSM4_FAGSY	Q9zru70 fagus sylla
266	27	100.0	129	2	Q8LAC6_ARATH	Q8lac6 arabidopsis	339	27	100.0	148	2	Q6IHD5_DROME	Q6ihd5 drosophila
267	27	100.0	129	2	Q8RWJ9_ARATH	Q8rwj9 arabidopsis	340	27	100.0	148	2	Q69SM7_ORYSA	Q69sm7 oryza sativ
268	27	100.0	129	2	Q6N719_RHOPA	Q6n719 rhodopsendo	341	27	100.0	148	2	Q6YRL5_ORYSA	Q6yrl5 oryza sativ
269	27	100.0	129	2	Q6PZC7_BAQU	Q6pzc7 bartonella	342	27	100.0	148	2	Q7XN05_ORYSA	Q7xn05 oryza sativ
270	27	100.0	129	2	Q6GX20_BARHE	Q6gx20 bartonella	343	27	100.0	148	2	Q84VE6_ORYSA	Q84ve6 oryza sativ
271	27	100.0	129	2	Q6NSP9_MOUSE	Q6nsp9 mus musculus	344	27	100.0	149	2	Q7XU68_ORYSA	Q7xu68 oryza sativ
272	27	100.0	129	2	Q76DT4_ONCMA	Q76dt4 oncorhynch	345	27	100.0	149	2	Q84T29_ORYSA	Q84t29 oryza sativ
273	27	100.0	130	2	Q7S9L9_NEUCR	Q7s9l9 neurospora	346	27	100.0	150	2	Q8MX16_9CARA	Q8mx16 arthropteru
274	27	100.0	130	2	Q4X2L7_PLACH	Q4x2l7 plasmodium	347	27	100.0	151	1	SMD3_DROME	Q4437 drosophila
275	27	100.0	130	2	Q82DY9_STRAW	Q82dy9 streptomyce	348	27	100.0	151	2	Q9HLX5_THEAC	Q9hly5 thermoplaem
276	27	100.0	130	2	Q8QP11_HPBVO	Q8qp11 hepatitis b	349	27	100.0	151	2	Q6ZT06_HUMAN	Q6zt06 homo sapien
277	27	100.0	132	2	Q6Z213_ORYSA	Q6z213 oryza sativ	350	27	100.0	151	2	Q6K6N6_ORYSA	Q6k6n6 oryza sativ
278	27	100.0	133	2	Q7SGW6_NEUCR	Q7sgw6 neurospora	351	27	100.0	151	2	Q6YX17_ORYSA	Q6yx17 oryza sativ
279	27	100.0	133	2	Q69KD3_ORYSA	Q69kd3 oryza sativ	352	27	100.0	151	2	Q4RB98_TETNG	Q4rb98 tetraodon n
280	27	100.0	133	2	Q5ZCM6_ORYSA	Q5zcm6 oryza sativ	353	27	100.0	152	1	HMGYB_SOYEN	Q10370 glycine max
281	27	100.0	134	2	Q5DGT7_SCHUA	Q5dgt7 schistosoma	354	27	100.0	152	2	Q84SV2_ORYSA	Q84sv2 oryza sativ
282	27	100.0	134	2	Q8W6J8_9CAUD	Q8w6j8 sinorhizobia	355	27	100.0	152	2	Q6DGQ4_BRARE	Q6dgq4 brachydanio
283	27	100.0	135	2	Q9HA29_HUMAN	Q9ha29 homo sapien	356	27	100.0	153	2	Q8NFD4_HUMAN	Q8nfd4 homo sapien
284	27	100.0	135	2	Q5CVY2_CRYPV	Q5cvy2 cryptospori	357	27	100.0	153	2	Q5JMM4_ORYSA	Q5jmm4 oryza sativ
285	27	100.0	135	2	Q6VYH4_ORYSA	Q6vyh4 oryza sativ	358	27	100.0	153	2	Q6A663_PROAC	Q6a663 propionibac
286	27	100.0	135	2	Q76DT3_ONCMA	Q76dt3 oncorhynch	359	27	100.0	154	2	Q41BT8_GIBZE	Q41bt8 gibberella
287	27	100.0	136	1	SMD3_CAEEL	Q17348 caenorhabdi	360	27	100.0	154	2	Q6PH81_HUMAN	Q6ph81 homo sapien
288	27	100.0	136	2	Q4PHV0_USTMA	Q4phv0 ustilago ma	361	27	100.0	154	2	Q9N4M2_CAEEL	Q9n4m2 caenorhabdi
289	27	100.0	136	2	Q7QPE9_GIALA	Q7qpe9 giardia lam	362	27	100.0	154	2	Q49512_ARATH	Q49512 arabidopsis
290	27	100.0	136	2	Q62BD6_CABER	Q62bd6 caenorhabdi	363	27	100.0	154	2	Q9CR55_MOUSE	Q9cr55 m mus muscu
291	27	100.0	136	2	Q61JY0_DROME	Q61jy0 drosophila	364	27	100.0	155	2	Q7SGK5_NEUCR	Q7sgk5 neurospora
292	27	100.0	136	2	Q6H583_ORYSA	Q6h583 oryza sativ	365	27	100.0	155	2	Q61KF8_DROME	Q61kf8 drosophila
293	27	100.0	137	2	Q96Z26_SULTO	Q96z26 sulfolobus	366	27	100.0	156	1	PSFN_MOUSE	Q70300 mus musculu
294	27	100.0	137	2	Q96D13_HUMAN	Q96d13 homo sapien	367	27	100.0	156	2	Q67VW7_ORYSA	Q67vw7 oryza sativ
295	27	100.0	137	2	Q22204_CAEEL	Q22204 caenorhabdi	368	27	100.0	157	2	Q6ZDG7_ORYSA	Q6zdg7 oryza sativ
296	27	100.0	137	2	Q6K6T3_ORYSA	Q6k6t3 oryza sativ	369	27	100.0	157	2	Q6ZKY5_ORYSA	Q6zky5 oryza sativ
297	27	100.0	137	2	Q5JKB2_ORYSA	Q5jkb2 oryza sativ	370	27	100.0	158	2	Q42927_SCHPO	Q42927 schizosacch
298	27	100.0	137	2	Q5SPH8_BRARE	Q5sph8 brachydanio	371	27	100.0	159	1	RL22_THEMA	P38511 thermotoga
299	27	100.0	138	2	Q6CHU2_YARLI	Q6chu2 yarrowia li	372	27	100.0	159	2	Q67U05_ORYSA	Q67u05 oryza sativ
300	27	100.0	138	2	Q6ZDL6_ORYSA	Q6zdl6 oryza sativ	373	27	100.0	159	2	Q4NZH4_9DELT	Q4nzh4 anaeromyxob
301	27	100.0	138	2	Q5ZAE9_ORYSA	Q5zae9 oryza sativ	374	27	100.0	159	2	Q4SUD2_TETNG	Q4sud2 tetraodon n
302	27	100.0	138	2	Q7NSS3_CHRVO	Q7nsa3 chromobacte	375	27	100.0	160	2	Q522L3_MAGGR	Q522l3 magnaporthe
303	27	100.0	138	2	Q6GL15_XENTR	Q6gl15 xenopus tro	376	27	100.0	160	2	Q52C81_ORYSA	Q52c81 oryza sativ
304	27	100.0	139	1	IGF_MYXGL	F22618 myxine gluc	377	27	100.0	160	2	Q6EEG6_LEIXC	Q6eeg6 leifsonia x
305	27	100.0	139	2	Q75177_ORYSA	Q75177 oryza sativ	378	27	100.0	161	2	Q6ZVT8_HUMAN	Q6zvt8 homo sapien
306	27	100.0	139	2	Q90W70_ONCMY	Q90w70 oncorhynch	379	27	100.0	161	2	Q5TSN9_ANOGA	Q5tsn9 anopheles g
307	27	100.0	140	2	Q8SIC2_ORYSA	Q8sic2 oryza sativ	380	27	100.0	161	2	Q5N7F3_ORYSA	Q5n7f3 oryza sativ
308	27	100.0	140	2	Q8B0B0_MOUSE	Q8b0b0 mus musculus	381	27	100.0	161	2	Q67M21_SYMTM	Q67m21 symbiobacte
309	27	100.0	141	2	Q7E2L9_ORYSA	Q7ezl9 oryza sativ	382	27	100.0	161	2	Q6PGX7_BRARE	Q6pgx7 brachydanio
310	27	100.0	141	2	Q5HAF6_GLUOX	Q5hxf6 gluconobact	383	27	100.0	162	2	Q81VRL_HUMAN	Q81vr1 homo sapien
311	27	100.0	141	2	Q9CZ23_MOUSE	Q9cz23 mus musculus	384	27	100.0	162	2	Q8SX06_DROME	Q8sx06 drosophila
312	27	100.0	141	2	Q4TIJ1_TETNG	Q4tij1 tetraodon n	385	27	100.0	162	2	Q5VME4_ORYSA	Q5vme4 oryza sativ
313	27	100.0	142	1	TOLR_ECOLI	P05829 escherichia	386	27	100.0	162	2	Q69KU8_ORYSA	Q69ku8 oryza sativ
314	27	100.0	142	1	Q9H7X2_HUMAN	Q9h7x2 homo sapien	387	27	100.0	162	2	Q53KT0_ORYSA	Q53kt0 oryza sativ
315	27	100.0	142	2	Q4QB16_LEIMA	Q4qb16 leishmania	388	27	100.0	163	2	Q59YN8_CANAL	Q59yn8 candida alb
316	27	100.0	142	2	Q6ZCF1_ORYSA	Q6zcf1 oryza sativ	389	27	100.0	163	2	Q59YD7_CANAL	Q59yd7 candida alb
317	27	100.0	142	2	Q6ZGV1_ORYSA	Q6zgv1 oryza sativ	390	27	100.0	163	2	Q5N720_ORYSA	Q5n720 oryza sativ
318	27	100.0	142	2	Q57RK5_SALCH	Q57rk5 salmonella	391	27	100.0	163	2	Q5ZAV3_ORYSA	Q5zav3 oryza sativ
319	27	100.0	142	2	Q5FTY8_GLUOX	Q5fty8 gluconobact	392	27	100.0	163	2	Q9JLY9_MOUSE	Q9jly9 mus musculu
320	27	100.0	142	2	Q5PM25_SALPA	Q5pm25 salmonobact	393	27	100.0	163	2	Q5ERC8_CARAU	Q5erc8 carassius a
321	27	100.0	142	2	Q7CQX0_SALTY	Q7cqxo salmonella	394	27	100.0	163	2	Q801Z9_BRARE	Q801z9 brachydanio
322	27	100.0	142	2	Q8XF53_SALTI	Q8xf53 salmonella	395	27	100.0	164	2	Q9AUR6_ORYSA	Q9aur6 oryza sativ
323	27	100.0	143	2	Q7Q0C0_ANOGA	Q7q0c0 anopheles g	396	27	100.0	165	2	Q416L4_GIBZE	Q416l4 gibberella

397	27	100.0	165	2	Q6UAZ0_9VIRU	Q6UAZ0 ralstonia p	470	27	100.0	178	2	Q42492_CANGL	Q42492 canavalia g
398	27	100.0	165	2	Q6ESI1_ORYSA	Q6esi1 oryza sativ	471	27	100.0	178	2	P70000_XENLA	P70000 xenopus lae
399	27	100.0	165	2	Q6ZDU0_ORYSA	Q6zdu0 oryza sativ	472	27	100.0	178	2	Q9W6Q5_BUFAR	Q9w6q5 bufo arenar
400	27	100.0	165	2	Q7XG68_ORYSA	Q7xg68 oryza sativ	473	27	100.0	179	2	Q52CE7_MAGGR	Q52ce7 magnaporthe
401	27	100.0	165	2	Q8RZ21_ORYSA	Q8rzz1 oryza sativ	474	27	100.0	179	2	Q9UKB0_HUMAN	Q9ukb0 homo sapien
402	27	100.0	165	2	Q8S6W9_ORYSA	Q8s6w9 oryza sativ	475	27	100.0	179	2	Q5YVGS_NOCFA	Q5yvg5 nocardia fa
403	27	100.0	165	2	Q5ZKT4_CHICK	Q5zkt4 gallus gall	476	27	100.0	179	2	Q8C340_MOUSE	Q8c340 mus musculus
404	27	100.0	165	2	Q9IB77_BRARE	Q9ib77 brachydanio	477	27	100.0	180	2	Q676V8_GASPA	Q676v8 hyacinthus
405	27	100.0	166	1	YKG3_CAEEL	P46553 caenorhabdi	478	27	100.0	180	2	Q9KZM9_STRCO	Q9kzm9 streptomyce
406	27	100.0	166	2	O143Z7_SCHPO	O14327 schizosacch	479	27	100.0	181	2	Q4XQW9_ASFFU	Q4xow9 aspergillus
407	27	100.0	166	2	Q6LHX3_DROME	Q6lhx3 drosophila	480	27	100.0	181	2	Q6P190_HUMAN	Q6p190 homo sapien
408	27	100.0	166	2	Q84SL3_ORYSA	Q84sl3 oryza sativ	481	27	100.0	181	2	Q9AXF8_CHLRE	Q9axf8 chlamydomon
409	27	100.0	166	2	Q8W5C0_ORYSA	Q8w5c0 oryza sativ	482	27	100.0	181	2	Q8BRP6_MOUSE	Q8brp6 mus musculus
410	27	100.0	166	2	O52559_AMYMD	O52559 amycolatops	483	27	100.0	182	2	Q6LBT6_HUMAN	Q6lbt6 homo sapien
411	27	100.0	167	2	O415C7_GIBZE	O415c7 gibberella	484	27	100.0	182	2	Q5VU81_HUMAN	Q5vu81 homo sapien
412	27	100.0	167	2	Q5CY21_CRPV	Q5cy21 cryptospori	485	27	100.0	182	2	Q518M7_ENTHI	Q518m7 entamoeba h
413	27	100.0	167	2	Q5CMG8_CRYHO	Q5cmg8 cryptospori	486	27	100.0	182	2	Q852D4_ORYSA	Q852d4 oryza sativ
414	27	100.0	167	2	Q5Z4W8_ORYSA	Q5z4w8 oryza sativ	487	27	100.0	182	2	Q7NK06_GLOVI	Q7nk06 gloeobacter
415	27	100.0	167	2	Q7EY20_ORYSA	Q7ey20 oryza sativ	488	27	100.0	183	2	O74333_SCHPO	O74333 schizosacch
416	27	100.0	167	2	Q4RJS7_TETNG	Q4rjs7 tetraodon n	489	27	100.0	183	2	O6TFE9_CAEPTA	O6tfe9 caedibacter
417	27	100.0	168	2	O7Q1E3_ANOGA	O7q1e3 anopheles g	490	27	100.0	184	2	O7CXH2_AGR75	O7cxh2 agrobactari
418	27	100.0	168	2	Q4Q1X7_LEIMA	Q4q1x7 leishmania	491	27	100.0	184	2	O7YWG2_NOCFA	O7ywg2 nocardia fa
419	27	100.0	168	2	Q91594_XENLA	Q91594 xenopus lae	492	27	100.0	185	2	Q8GT10_ORYSA	Q8gt10 oryza sativ
420	27	100.0	170	1	FGF2_CAVPO	Q60487 cavia porce	493	27	100.0	185	2	Q5Z256_NOCFA	Q5z256 nocardia fa
421	27	100.0	170	2	Q38778_AVESEA	Q38778 avena sativ	494	27	100.0	185	2	Q62BW2_BURMA	Q62bw2 burkholderi
422	27	100.0	170	2	Q94LT6_ORYSA	Q94lt6 oryza sativ	495	27	100.0	185	2	Q8BCA9_HPBVO	Q8bca9 hepatitis b
423	27	100.0	170	2	O6VMJ4_ORYSA	O6vmj4 oryza sativ	496	27	100.0	186	2	O53K81_ORYSA	O53k81 oryza sativ
424	27	100.0	170	2	Q6AUV4_ORYSA	Q6auv4 oryza sativ	497	27	100.0	186	2	Q3A6R0_CAUCR	Q3a6r0 caulobacter
425	27	100.0	170	2	Q9D590_MOUSE	Q9d590 mus musculus	498	27	100.0	187	2	Q5BDF0_EMENI	Q5bdf0 aspergillus
426	27	100.0	170	2	Q91A81_BRARE	Q91a81 brachydanio	499	27	100.0	188	2	Q6Z610_ORYSA	Q6z610 oryza sativ
427	27	100.0	171	2	Q8UD59_AGR75	Q8ud59 agrobactari	500	27	100.0	189	1	NLAL2_ARATH	Q9fk53 arabidopsis
428	27	100.0	172	1	CRBP_HUMAN	O14011 homo sapien	501	27	100.0	189	2	O71E78_HUMAN	O71e78 homo sapien
429	27	100.0	172	2	O53XX5_HUMAN	O53xx5 homo sapien	502	27	100.0	189	2	O5NA66_ORYSA	O5na66 oryza sativ
430	27	100.0	172	2	Q4RSL7_MACFA	Q4rsl7 macaca fasc	503	27	100.0	189	2	O75H67_ORYSA	Q75h67 oryza sativ
431	27	100.0	172	2	Q5RFL3_PONPY	Q5rfl3 pongo pygma	504	27	100.0	189	2	Q8LRU5_WHEAT	Q8lrus triticum ae
432	27	100.0	172	2	Q6K6S0_ORYSA	Q6k6s0 oryza sativ	505	27	100.0	189	2	Q84G07_FARPEN	Q84g07 paracoccu
433	27	100.0	172	2	Q75V41_PHPYA	Q75v41 physcomitre	506	27	100.0	190	2	Q4P2V8_USTMA	Q4p2v8 ustilago ma
434	27	100.0	173	2	Q4WAU6_ASFFU	Q4wau6 aspergillus	507	27	100.0	190	2	Q6BWR1_DEBHA	Q6bwr1 debaryomyce
435	27	100.0	173	2	Q5P5X7_AZOSE	Q5p5x7 azoarcus sp	508	27	100.0	190	2	Q86RU0_PDIOF	Q86ru0 teloglabrus
436	27	100.0	173	2	Q4TGS1_TETNG	Q4tgs1 tetraodon n	509	27	100.0	190	2	Q9S7Z7_MAIZE	Q9s7z7 zea mays (m
437	27	100.0	174	2	Q524C8_MAGGR	Q524c8 magnaporthe	510	27	100.0	190	2	O5ZL08_CHICK	Q5zlu8 gallus gall
438	27	100.0	174	2	Q52D13_MAGGR	Q52d13 magnaporthe	511	27	100.0	191	2	O5S5S6_CRYNE	Q5s5s6 cryptococcu
439	27	100.0	174	2	Q4IMP1_GIBZE	Q4imp1 gibberella	512	27	100.0	191	2	O46029_CHITE	O46029 chironomu
440	27	100.0	174	2	Q69PA0_ORYSA	Q69pa0 oryza sativ	513	27	100.0	191	2	O5DEX9_SCHJA	O5dex9 schistosoma
441	27	100.0	174	2	O7XFY4_ORYSA	O7xfy4 oryza sativ	514	27	100.0	191	2	O6YX52_ORYSA	O6yx52 oryza sativ
442	27	100.0	174	2	Q8S831_ORYSA	Q8s831 oryza sativ	515	27	100.0	191	2	Q7XHU1_QUEERO	Q7xhj1 quercus rob
443	27	100.0	174	2	Q94E12_ORYSA	Q94e12 oryza sativ	516	27	100.0	191	2	Q9FER8_MAIZE	Q9fer8 zea mays (m
444	27	100.0	174	2	Q6X9P4_9VIRU	Q6x9p4 macrobrachi	517	27	100.0	192	2	O4P9S1_USTMA	Q4p9s1 ustilago ma
445	27	100.0	175	2	Q51292_MAGGR	Q51292 magnaporthe	518	27	100.0	192	2	O5D1L0_PDIOF	Q5d1l0 sphyracepha
446	27	100.0	175	2	Q3DE71_SCHJA	Q3de71 schistosoma	519	27	100.0	192	2	Q8H2Y1_ORYSA	Q8h2y1 oryza sativ
447	27	100.0	175	2	Q854K2_ORYSA	Q854k2 oryza sativ	520	27	100.0	192	2	Q8Y248_PAGMA	Q8y248 pagrus majo
448	27	100.0	175	2	Q544Z5_ORYSA	Q544z5 oryza sativ	521	27	100.0	193	2	Q84TW4_ORYSA	Q84tw4 oryza sativ
449	27	100.0	175	2	Q544Z5_ORYSA	Q544z5 oryza sativ	522	27	100.0	193	2	Q9FED9_ORYSA	Q9fed9 oryza sativ
450	27	100.0	175	2	Q6Z6A1_ORYSA	Q6z6a1 oryza sativ	523	27	100.0	193	2	Q9M994_ARATH	Q9m994 arabidopsis
451	27	100.0	175	2	Q5P8W2_AZOSE	Q5p8w2 azoarcus sp	524	27	100.0	193	2	O9FYSS_MAIZE	O9fys5 zea mays (m
452	27	100.0	176	1	HMGYA_SOYBN	Q66cc3 xiphophorus	525	27	100.0	193	2	Q9KWH6_STRCO	Q9kwh6 streptomyce
453	27	100.0	176	2	Q84YP5_ORYSA	Q84yp5 oryza sativ	526	27	100.0	194	1	3MCH_AERPE	Q9y9p1 aeropyrum p
454	27	100.0	176	2	Q3WKK2_ORYSA	Q3wkk2 oryza sativ	527	27	100.0	194	2	Q4HLD5_LEIMA	Q4hld5 leishmania
455	27	100.0	176	2	Q94E12_ORYSA	Q94e12 oryza sativ	528	27	100.0	194	2	Q84NR9_ORYSA	Q84nr9 oryza sativ
456	27	100.0	176	2	O7WXL6_ALCEU	O7wxl6 alcaligenes	529	27	100.0	194	2	Q8W2U0_ORYSA	Q8w2u0 oryza sativ
457	27	100.0	177	2	O50091_PVRHO	O50091 pyrococcus	530	27	100.0	194	2	Q9AHP3_9ACTO	Q9ahp3 arcanobacte
458	27	100.0	177	2	Q54KX4_DICDI	Q54kx4 dictyosteli	531	27	100.0	194	2	O9HXN0_PSEAE	O9hxn0 pseudomonas
459	27	100.0	177	2	Q4Q8I4_LEIMA	Q4q8i4 leishmania	532	27	100.0	195	2	Q4IQN4_GIBZE	Q4iqn4 gibberella
460	27	100.0	177	2	Q8N7U3_ORYSA	Q8n7u3 oryza sativ	533	27	100.0	195	2	Q6EP78_ORYSA	Q6ep78 oryza sativ
461	27	100.0	177	2	Q5QN35_ORYSA	Q5qn35 oryza sativ	534	27	100.0	195	2	O53LD1_ORYSA	Q53ld1 oryza sativ
462	27	100.0	177	2	Q5SMQ7_ORYSA	Q5smq7 oryza sativ	535	27	100.0	196	2	P78443_HUMAN	P78443 homo sapien
463	27	100.0	177	2	Q528B3_ORYSA	Q528b3 oryza sativ	536	27	100.0	196	2	Q84S46_ORYSA	Q84s46 oryza sativ
464	27	100.0	177	2	Q5Z829_ORYSA	Q5z829 oryza sativ	537	27	100.0	196	2	O5JMF5_ORYSA	O5jmr5 oryza sativ
465	27	100.0	177	2	Q5Z3N9_NOCFA	Q5z3n9 nocardia fa	538	27	100.0	196	2	O53L29_ORYSA	O53l29 oryza sativ
466	27	100.0	178	2	Q8H3M0_ORYSA	Q8h3m0 oryza sativ	539	27	100.0	196	2	Q4TG7_TETNG	Q4tgy7 tetraodon n
467	27	100.0	178	2	Q6ZJ93_ORYSA	Q6zj93 oryza sativ	540	27	100.0	197	2	Q4H3Q9_CIOIN	Q4h3q9 ciona intes
468	27	100.0	178	2	O7XTX4_ORYSA	O7xtx4 oryza sativ	541	27	100.0	197	2	Q43877_PEA	Q43877 pisum sativ
469	27	100.0	178	2	Q42461_CANGL	Q42461 canavalia g	542	27	100.0	197	2	Q4RBG5_TETNG	Q4rbg5 tetraodon n

543	27	100.0	198	2	Q9H8H4_HUMAN	Q9h8h4 homo sapien	616	27	100.0	206	2	Q9TBS1_THEVO	Q97be1 thermoplaem
544	27	100.0	198	2	Q84T32_ORYSA	Q84t32 oryza sativ	617	27	100.0	206	2	Q5WA80_ORYSA	Q5wa80 oryza sativ
545	27	100.0	198	2	Q7XY12_CHLS6	Q7xy12 chlorarachn	618	27	100.0	206	2	Q6ZAF4_ORYSA	Q6zaf4 oryza sativ
546	27	100.0	198	2	Q6AHG0_LEIFX6	Q6ahg0 leifsonia x	619	27	100.0	206	2	Q7XB88_ORYSA	Q7xb88 oryza sativ
547	27	100.0	198	2	Q72H15_THET2	Q72h15 thermus the	620	27	100.0	206	2	Q4H9E1_9DEIO	Q4h9e1 deinococcus
548	27	100.0	198	2	Q89WU2_BRAJA	Q89wu2 bradyrhizob	621	27	100.0	206	2	Q9BK47_STRCO	Q9bk47 streptomyce
549	27	100.0	198	2	Q69469_9BETA	Q69469 human herpe	622	27	100.0	206	2	Q4V2N7_BURMA	Q4v2n7 burkholderi
550	27	100.0	198	2	Q8V5Y3_9ALPH	Q8v5y3 gallid herp	623	27	100.0	206	2	Q7ZWN7_XENLA	Q7zwn7 xenopus lae
551	27	100.0	198	2	Q9DGS3_MEHV1	Q9dgs3 meleagrid h	624	27	100.0	207	2	Q5W6D3_ORYSA	Q5w6d3 oryza sativ
552	27	100.0	198	2	Q76QS7_9BETA	Q76qs7 human herpe	625	27	100.0	207	2	Q9D1W7_MOUSE	Q9d1w7 mus musculu
553	27	100.0	198	2	Q5TYZ7_BRARE	Q5tyz7 brachydanio	626	27	100.0	208	2	Q529I3_MAGGR	Q529i3 magnaporthe
554	27	100.0	198	2	Q5VN72_ORYSA	Q5vn72 oryza sativ	627	27	100.0	208	2	Q4HXN7_GIBZE	Q4hxn7 gibberella
555	27	100.0	199	2	Q6ZGS1_ORYSA	Q6zgs1 oryza sativ	628	27	100.0	208	2	Q622I3_CABEL	Q622i3 caenorhabdi
556	27	100.0	199	2	Q9LYB2_ARATH	Q9lyb2 arabidopsis	629	27	100.0	208	2	Q5IST1_MACFA	Q5ist1 macaca fasc
557	27	100.0	199	2	Q4RDZ2_TETNG	Q4rdz2 tetraodon n	630	27	100.0	208	2	Q82964_RALSO	Q82964 ralstonia s
558	27	100.0	200	2	Q86RZ2_9DIOP	Q86rz2 sphyracepha	631	27	100.0	208	2	Q8KQM0_SACER	Q8kqm0 saccharopol
559	27	100.0	200	2	Q67IV9_ORYSA	Q67iv9 oryza sativ	632	27	100.0	208	2	Q4NN51_9DELT	Q4nn51 anaeromyxob
560	27	100.0	201	2	Q86RY1_9DIOP	Q86ry1 diasemopsis	633	27	100.0	208	2	Q4NUC7_9DELT	Q4nuc7 anaeromyxob
561	27	100.0	201	2	Q86RY2_9DIOP	Q86ry2 diasemopsis	634	27	100.0	208	2	Q62266_PAGWA	Q62266 pagrus majo
562	27	100.0	201	2	Q86RZ4_9DIOP	Q86rz4 diasemopsis	635	27	100.0	209	2	Q8H5G7_ORYSA	Q8h5g7 oryza sativ
563	27	100.0	201	2	Q33374_NEIGO	Q33374 neisseria g	636	27	100.0	209	2	Q5NZS7_AZOSE	Q5nzs7 azoarcus sp
564	27	100.0	201	2	Q5F510_NEIG1	Q5f510 neisseria g	637	27	100.0	209	2	Q76IC5_RAT	Q76ic5 rattus norv
565	27	100.0	201	2	Q9JW85_NEIMA	Q9jw85 neisseria m	638	27	100.0	210	1	Y2543_HALSA	Q9hnh2 halobacteri
566	27	100.0	201	2	Q9JXN8_NEIMB	Q9jxn8 neisseria m	639	27	100.0	210	1	Q5KLP3_CRYNE	Q5klp3 cryptococcu
567	27	100.0	202	1	NLALI1_ARATH	Q8vzt0 arabidopsis	640	27	100.0	210	2	Q5TY95_CRYNE	Q5ty95 cryptococcu
568	27	100.0	202	2	Q29683_ARCFU	Q29683 archaeoglob	641	27	100.0	210	2	Q7KZ72_HUMAN	Q7kz72 homo sapien
569	27	100.0	202	2	Q5T7Y9_HUMAN	Q5t7y9 homo sapien	642	27	100.0	210	2	Q5DEQ2_SCHJA	Q5deq2 schistosoma
570	27	100.0	202	2	Q95NZ5_SCHJA	Q95nz5 schistosoma	643	27	100.0	210	2	Q4LLI0_9BURK	Q4lli0 burkholderi
571	27	100.0	202	2	Q612S3_CABER	Q61zs3 caenorhabdi	644	27	100.0	210	2	Q9RVL7_DEIRA	Q9rvl7 deinococcus
572	27	100.0	202	2	Q9LLI8_STRCO	Q9lli8 streptomyce	645	27	100.0	210	2	Q92IK4_MOUSE	Q92ik4 mus musculu
573	27	100.0	202	2	Q66H75_RAT	Q66h75 rattus norv	646	27	100.0	210	2	Q695I1_9BETA	Q695i1 human herpe
574	27	100.0	203	2	Q8NI61_HUMAN	Q8ni61 homo sapien	647	27	100.0	210	2	Q9IC81_9VIRU	Q9ic81 human astro
575	27	100.0	203	2	Q86RY5_9DIOP	Q86ry5 sphyracepha	648	27	100.0	210	2	Q9IC85_9VIRU	Q9ic85 human astro
576	27	100.0	203	2	Q86RY9_9DIOP	Q86ry9 sphyracepha	649	27	100.0	210	2	Q6VTL9_NPVCD	Q6vtl9 choristoneu
577	27	100.0	203	2	Q18036_CABEL	Q18036 caenorhabdi	650	27	100.0	211	1	MDCG_PSESM	Q87v59 pseudomonas
578	27	100.0	203	2	Q84QZ2_ORYSA	Q84qz2 oryza sativ	651	27	100.0	211	2	Q4NI15_ASFPU	Q4ni15 aspergillus
579	27	100.0	203	2	Q8H8F1_ORYSA	Q8h8f1 oryza sativ	652	27	100.0	211	2	Q4WAK1_ASFPU	Q4wxk1 aspergillus
580	27	100.0	203	2	Q9SEH2_BRANA	Q9seh2 brassica na	653	27	100.0	211	2	Q6EPP9_ORYSA	Q6ep9 oryza sativ
581	27	100.0	203	2	Q7VVM2_BORPE	Q7vvm2 bordetella	654	27	100.0	211	2	Q94GR2_ORYSA	Q94gr2 oryza sativ
582	27	100.0	203	2	Q7WAV8_BORPA	Q7wav8 bordetella	655	27	100.0	211	2	Q9SX47_ARATH	Q9sx47 arabidopsis
583	27	100.0	203	2	Q7WK15_BORBR	Q7wk15 bordetella	656	27	100.0	211	2	Q4NQ03_9DELT	Q4nq03 anaeromyxob
584	27	100.0	203	2	Q931I3_9BETA	Q931i3 human herpe	657	27	100.0	212	2	Q52DT2_MAGGR	Q52dt2 magnaporthe
585	27	100.0	203	2	Q4TFH2_TETNG	Q4tfh2 tetraodon n	658	27	100.0	212	2	Q94AD1_ARATH	Q94ad1 arabidopsis
586	27	100.0	204	1	BTf3_MOUSE	Q64152 mus musculu	659	27	100.0	212	2	Q5YTN2_NOCPA	Q5ynt2 nocardia fa
587	27	100.0	204	2	Q4PBW2_USTWA	Q4pbw2 ustilago ma	660	27	100.0	212	2	Q5UIX5_RAT	Q5uix5 rattus norv
588	27	100.0	204	2	Q7XTX9_ORYSA	Q7txx9 oryza sativ	661	27	100.0	212	2	Q917H4_HPBO	Q917h4 hepatis b
589	27	100.0	204	2	P92954_ARATH	P92954 arabidopsis	662	27	100.0	212	2	Q4FDF2_HPBO	Q4fdf2 hepatis b
590	27	100.0	204	2	Q43386_ARATH	Q43386 arabidopsis	663	27	100.0	213	2	Q52CY8_MAGGR	Q52cy8 magnaporthe
591	27	100.0	204	2	Q6XZR1_9RETR	Q6xxr1 simian foam	664	27	100.0	213	2	Q5DFE6_SCHJA	Q5df6 schistosoma
592	27	100.0	204	2	Q6XZR2_9RETR	Q6xxr2 simian foam	665	27	100.0	213	2	Q5D8H8_SCHJA	Q5d8h8 schistosoma
593	27	100.0	204	2	Q6XZR3_9RETR	Q6xxr3 simian foam	666	27	100.0	213	2	Q69PJ0_ORYSA	Q69pj0 oryza sativ
594	27	100.0	204	2	Q6XZR4_9RETR	Q6xxr4 simian foam	667	27	100.0	213	2	Q43600_ORYSA	Q43600 oryza sativ
595	27	100.0	204	2	Q6XZR5_9RETR	Q6xxr5 simian foam	668	27	100.0	213	2	Q69MW7_ORYSA	Q69mw7 oryza sativ
596	27	100.0	204	2	Q6XZR6_9RETR	Q6xxr6 simian foam	669	27	100.0	213	2	Q4ZZA5_PSESY	Q4zza5 pseudomonas
597	27	100.0	204	2	Q6XZR7_9RETR	Q6xxr7 simian foam	670	27	100.0	214	2	Q6XZR9_9RETR	Q6xxr9 simian foam
598	27	100.0	204	2	Q6XZS0_9RETR	Q6xzso simian foam	671	27	100.0	214	2	Q52IK9_MAGGR	Q52ik9 magnaporthe
599	27	100.0	204	2	Q6XZS1_9RETR	Q6xzsl simian foam	672	27	100.0	214	2	Q585Y5_9TRYP	Q585y5 trypanosoma
600	27	100.0	204	2	Q6XZS2_9RETR	Q6xzsz simian foam	673	27	100.0	214	2	Q5ZCH1_ORYSA	Q5zch1 oryza sativ
601	27	100.0	204	2	Q6XZS3_9RETR	Q6xzss simian foam	674	27	100.0	214	2	Q6YRP0_ORYSA	Q6yrp0 oryza sativ
602	27	100.0	204	2	Q6XZS4_9RETR	Q6xzss4 simian foam	675	27	100.0	214	2	Q765C9_HPBO	Q765c9 hepatis b
603	27	100.0	204	2	Q6XZS5_9RETR	Q6xzss5 simian foam	676	27	100.0	214	2	Q8BCB0_HPBO	Q8bc80 hepatis b
604	27	100.0	204	2	Q6XZS6_9RETR	Q6xzss6 simian foam	677	27	100.0	215	2	Q4WTF4_ASFPU	Q4wtf4 aspergillus
605	27	100.0	204	2	Q6XZS7_9RETR	Q6xzss7 simian foam	678	27	100.0	215	2	Q6M905_NEUCR	Q6m905 neurospora
606	27	100.0	204	2	Q6XZS8_9RETR	Q6xzss8 simian foam	679	27	100.0	215	2	Q9BVX3_HUMAN	Q9bv3 homo sapien
607	27	100.0	204	2	Q6XZS9_9RETR	Q6xzss9 simian foam	680	27	100.0	215	2	Q5VMQ3_ORYSA	Q5vmq3 oryza sativ
608	27	100.0	204	2	Q6XZT0_9RETR	Q6xzt0 simian foam	681	27	100.0	215	2	Q5Z4N8_ORYSA	Q5z4n8 oryza sativ
609	27	100.0	205	2	Q523X2_MAGGR	Q523x2 magnaporthe	682	27	100.0	215	2	Q4SLJ6_TETNG	Q4slj6 tetraodon n
610	27	100.0	205	2	Q6C5N8_YARLI	Q6c5n8 yarrowia li	683	27	100.0	216	2	Q8XOV9_NEUCR	Q8xov9 neurospora
611	27	100.0	205	2	Q9YIN9_BRUMA	Q9yin9 brugia mala	684	27	100.0	216	2	Q4Q734_LEIMA	Q4q734 leishmania
612	27	100.0	205	2	Q67TZ5_ORYSA	Q67tz5 oryza sativ	685	27	100.0	216	2	Q4KD94_PSEF5	Q4kd94 pseudomonas
613	27	100.0	205	2	Q9I4E0_PSEAE	Q9i4e0 pseudomonas	686	27	100.0	217	2	Q6ZNR8_HUMAN	Q6znr8 homo sapien
614	27	100.0	206	1	BTf3_HUMAN	P20290 homo sapien	687	27	100.0	217	2	Q27277_SCHWA	Q27277 schistosoma
615	27	100.0	206	1	RHO3_SCHCO	Q9p8j9 schizophyll	688	27	100.0	217	2	Q583J0_9TRYP	Q583j0 trypanosoma

689	27	100.0	217	2	Q4YTX9_PLABE	Q4YTX9 plasmodium	762	27	100.0	234	2	Q6JMW4_9UROC	Q6jmw4 oikopleura
690	27	100.0	217	2	Q94J84_ORYSA	Q94j84 oryza sativ	763	27	100.0	234	2	Q851F3_ORYSA	Q851f3 oryza sativ
691	27	100.0	218	2	Q59F82_HUMAN	Q59f82 homo sapien	764	27	100.0	234	2	Q6Z5R6_ORYSA	Q6z5r6 oryza sativ
692	27	100.0	218	2	Q9BT43_HUMAN	Q9bt43 homo sapien	765	27	100.0	234	2	Q63VC8_BURPS	Q63vc8 burkholderi
693	27	100.0	218	2	Q8IM10_PLAF7	Q8im10 plasmodium	766	27	100.0	235	2	Q51Z70_MAGGR	Q51z70 magnaporthe
694	27	100.0	218	2	Q9D8G0_MOUSE	Q9d8g0 m mus muscu	767	27	100.0	235	2	Q8LMV9_ORYSA	Q8lmv9 oryza sativ
695	27	100.0	218	2	Q57165_9BROM	Q57165 spinach lat	768	27	100.0	235	2	Q9DC49_MOUSE	Q9dc49 mus musculus
696	27	100.0	218	2	Q661E8_BRARE	Q661e8 brachydanio	769	27	100.0	236	1	UL51_PRVKA	Q85227 pseudorabie
697	27	100.0	218	2	Q6TV21_XENLA	Q6tv21 xenopus lae	770	27	100.0	236	2	Q6AWX7_ORYSA	Q6awx7 oryza sativ
698	27	100.0	219	2	Q7RNY1_PLAYO	Q7rny1 plasmodium	771	27	100.0	236	2	Q6AWX7_ORYSA	Q6awx7 oryza sativ
699	27	100.0	219	2	Q6JAC7_TOBAC	Q6jac7 nicotiana t	772	27	100.0	236	2	Q5PP96_9ALPH	Q5pp96 suid herpes
700	27	100.0	219	2	Q7NEF0_GLOVI	Q7nef0 gloeobacter	773	27	100.0	237	2	Q4TBY7_TETNG	Q4tby7 tetraodon n
701	27	100.0	219	2	Q5XJG9_MOUSE	Q5xjg9 mus musculus	774	27	100.0	237	2	Q6XZR8_9RETR	Q6xvr7 oryza sativ
702	27	100.0	219	2	Q4TK73_TETNG	Q4tk73 tetraodon n	775	27	100.0	237	2	Q8REQ1_FUSNN	Q8reql simian foam
703	27	100.0	220	2	Q53NA0_ORYSA	Q53na0 oryza sativ	776	27	100.0	238	2	Q55WH0_CRYNE	Q55wh0 cryptococci
704	27	100.0	220	2	Q8BVG0_MOUSE	Q8bvg0 mus musculus	777	27	100.0	239	2	Q5DH51_SCHJA	Q5dh51 schistosoma
705	27	100.0	220	2	Q4SZX0_TETNG	Q4szx0 tetraodon n	778	27	100.0	239	2	Q3IBJ1_MEHVI	Q3ibj1 meleagrid h
706	27	100.0	221	2	Q8N7V5_HUMAN	Q8n7v5 homo sapien	779	27	100.0	240	1	ING5_HUMAN	Q8wyh8 homo sapien
707	27	100.0	221	2	Q5GQL6_9CAUD	Q5gql6 bacterioph	780	27	100.0	240	1	ING5_MOUSE	Q9d8y8 mus musculus
708	27	100.0	221	2	Q91H61_ARATH	Q91h61 arabidopsis	781	27	100.0	240	1	Q6ETX9_ORYSA	Q6etx9 oryza sativ
709	27	100.0	222	2	Q5VP19_ORYSA	Q5vp19 oryza sativ	782	27	100.0	240	2	Q6Z2R4_ORYSA	Q6z2r4 oryza sativ
710	27	100.0	222	2	Q6Z751_ORYSA	Q6z751 oryza sativ	783	27	100.0	240	2	Q9SQ54_SPIOL	Q9sq54 spinacia ol
711	27	100.0	222	2	Q69K36_ORYSA	Q69k36 oryza sativ	784	27	100.0	240	2	Q8C2X5_MOUSE	Q8czx5 mus musculus
712	27	100.0	222	2	Q7ZX70_XENLA	Q7zx70 xenopus lae	785	27	100.0	241	2	Q5QBH6_9DIPT	Q5qbh6 culicoides
713	27	100.0	223	1	RS5_GLOVI	Q7neq9 gloeobacter	786	27	100.0	241	2	Q7SLK8_ORYSA	Q7slk8 oryza sativ
714	27	100.0	223	2	Q4HXW7_GIBZE	Q4hxm7 gibberella	787	27	100.0	241	2	Q5M8K4_XENTR	Q5mhk4 xenopus tro
715	27	100.0	223	2	Q5T7Y8_HUMAN	Q5t7y8 homo sapien	788	27	100.0	241	2	Q6NV06_BRARE	Q6nv06 brachydanio
716	27	100.0	223	2	Q6R8J6_RABIT	Q6r8j6 cryctolagus	789	27	100.0	242	2	Q6EV07_9DYTI	Q6ev07 meladema co
717	27	100.0	223	2	Q7NG78_GLOVI	Q7ng78 gloeobacter	790	27	100.0	242	2	Q53PT8_ORYSA	Q53pt8 oryza sativ
718	27	100.0	223	2	Q4TAJ6_TETNG	Q4taj6 tetraodon n	791	27	100.0	242	2	Q8GFV9_CITFR	Q8gfv9 citrobacter
719	27	100.0	224	2	Q59VH0_CANAL	Q59vh0 candida alb	792	27	100.0	243	2	Q7Z334_HUMAN	Q7z334 homo sapien
720	27	100.0	224	2	Q5TVP6_ANOGA	Q5tvp6 anopheles g	793	27	100.0	243	2	Q9H7F9_HUMAN	Q9h7f9 homo sapien
721	27	100.0	224	2	Q6YSH6_ORYSA	Q6yeh6 oryza sativ	794	27	100.0	243	2	Q6QWA6_CAEBR	Q6qwa6 caenorhabdi
722	27	100.0	224	2	Q6VYQ3_ORYSA	Q6vyq3 oryza sativ	795	27	100.0	243	2	Q9C7H1_ARATH	Q9c7h1 arabidopsis
723	27	100.0	224	2	Q5SKB2_THET8	Q5skb2 thermus the	796	27	100.0	243	2	Q8C4X6_MOUSE	Q8c4x6 mus musculus
724	27	100.0	224	2	Q72KM3_THET2	Q72km3 thermus the	797	27	100.0	244	2	Q96S43_HUMAN	Q96s43 homo sapien
725	27	100.0	224	2	Q82T95_NITEU	Q82t95 nitrosomona	798	27	100.0	244	2	Q4TQ67_9SPHN	Q4tg67 erythrobaet
726	27	100.0	225	2	Q6BT88_DEBHA	Q6bt88 debaryomyce	799	27	100.0	245	2	Q5KEZ7_CRYNE	Q5kez7 cryptococci
727	27	100.0	225	2	Q8R0C0_MOUSE	Q8r0c0 mus musculus	800	27	100.0	245	2	Q9PUJ8_ALIMI	Q9puj8 alligator m
728	27	100.0	226	2	Q4NP13_9DELT	Q4np13 anaeromyxob	801	27	100.0	245	2	Q4RRM3_TETNG	Q4rrm3 tetraodon n
729	27	100.0	227	2	Q4P5P0_USTMA	Q4p5p0 ustilago ma	802	27	100.0	246	2	Q6TFY0_ERWAM	Q6tfy0 erwinia amy
730	27	100.0	227	2	Q5Z6L4_ORYSA	Q5z6l4 oryza sativ	803	27	100.0	246	2	Q4TLH8_TETNG	Q4tlh8 tetraodon n
731	27	100.0	227	2	Q7FAC3_ORYSA	Q7fac3 oryza sativ	804	27	100.0	247	2	Q9SQ53_SPIOL	Q9sq53 spinacia ol
732	27	100.0	227	2	Q7X6T6_ORYSA	Q7x6t6 oryza sativ	805	27	100.0	247	2	Q8GV23_CHLRE	Q8gv23 chlamydomon
733	27	100.0	227	2	Q8LG48_ARATH	Q8lg48 arabidopsis	806	27	100.0	248	1	HT31_HUMAN	Q9y3y2 homo sapien
734	27	100.0	227	2	Q93V14_ARATH	Q93v14 arabidopsis	807	27	100.0	248	2	Q5T7Z0_HUMAN	Q5t7z0 homo sapien
735	27	100.0	227	2	Q5N0M0_SYNP6	Q5n0m3 synecchococ	808	27	100.0	248	2	Q8MY63_CIOSA	Q8my63 ciona savig
736	27	100.0	227	2	Q4RZP7_TETNG	Q4rzp7 tetraodon n	809	27	100.0	248	2	Q4R6E5_MACFA	Q4r6e5 macaca fasc
737	27	100.0	228	2	Q6K6W3_ORYSA	Q6k6w3 oryza sativ	810	27	100.0	248	2	Q93501_XENLA	Q93501 xenopus lae
738	27	100.0	228	2	Q8LMC8_ORYSA	Q8lmc8 oryza sativ	811	27	100.0	249	1	HT31_MOUSE	Q9cy57 mus musculus
739	27	100.0	228	2	Q6LCS5_HPV20	Q6lcs5 human papil	812	27	100.0	249	2	Q8MUE9_9BIVA	Q8mue9 chlamya far
740	27	100.0	229	2	Q6A0Y5_9EURY	Q6a0y5 haloterrige	813	27	100.0	249	2	Q68VB6_NICBE	Q68vb6 nicotiana b
741	27	100.0	230	2	Q6TRP1_ORYSA	Q6trp1 oryza sativ	814	27	100.0	250	1	RS3_ACTAD	Q6t7r8 acinetobact
742	27	100.0	230	2	Q651G6_ORYSA	Q651g6 oryza sativ	815	27	100.0	250	2	Q5DET1_SCHJA	Q5det1 schistosoma
743	27	100.0	230	2	Q6W3L8_9PROT	Q6w3l8 alvinella p	816	27	100.0	250	2	Q6K1X1_ORYSA	Q6k1x1 oryza sativ
744	27	100.0	230	2	Q995P5_9INFA	Q995p5 influenza a	817	27	100.0	250	2	Q4RX21_TETNG	Q4rx21 tetraodon n
745	27	100.0	231	2	Q7Q5I0_GIALA	Q7q5i0 giardia lam	818	27	100.0	251	2	Q5VQV6_ORYSA	Q5vqv6 oryza sativ
746	27	100.0	231	2	Q6BPA8_ORYSA	Q6bpa8 oryza sativ	819	27	100.0	251	2	Q7F1F6_ORYSA	Q7f1f6 oryza sativ
747	27	100.0	231	2	Q9LQC2_ARATH	Q9lqc2 arabidopsis	820	27	100.0	251	2	Q8LM18_ORYSA	Q8lm18 oryza sativ
748	27	100.0	231	2	Q7ZW44_BRARE	Q7zw44 brachydanio	821	27	100.0	252	2	Q55Y92_CRYNE	Q55y92 cryptococci
749	27	100.0	232	2	Q4PFY6_USTMA	Q4pfy6 ustilago ma	822	27	100.0	252	2	Q75A81_ASHGO	Q75a81 ashbya goss
750	27	100.0	232	2	Q61PE2_HUMAN	Q61pe2 homo sapien	823	27	100.0	252	2	Q5KLP6_CRYNE	Q5klp6 cryptococci
751	27	100.0	232	2	Q9V9M3_DROME	Q9v9m3 drosophila	824	27	100.0	252	2	Q6X3V8_9VIRU	Q6x3v8 bacillus th
752	27	100.0	232	2	Q6VB67_HVIR	Q6vb67 human herpe	825	27	100.0	252	2	Q8Z0L7_ORYSA	Q8z0l7 oryza sativ
753	27	100.0	233	2	Q592V5_LYNST	Q592v5 lymanaea sta	826	27	100.0	252	2	Q68VB5_NICBE	Q68vb5 nicotiana b
754	27	100.0	233	2	Q6YIJ3_ORYSA	Q6yij3 oryza sativ	827	27	100.0	252	2	Q7WSF8_BACTU	Q7wsf8 bacillus th
755	27	100.0	233	2	Q7XF38_ORYSA	Q7xf38 oryza sativ	828	27	100.0	252	2	Q57K93_SALCH	Q57k93 salmonella
756	27	100.0	233	2	Q9AYG8_ORYSA	Q9ayg8 oryza sativ	829	27	100.0	252	2	Q8Z3Z5_SALTI	Q8z3z5 salmonella
757	27	100.0	233	2	Q6PWR1_9GAMM	Q6pwr1 uncultured	830	27	100.0	252	2	Q8ZM93_SALTY	Q8zm93 salmonella
758	27	100.0	233	2	Q6ZFH7_BURNA	Q6zfh7 burkholderi	831	27	100.0	252	2	Q5FEQ3_SALPA	Q5fed3 schizosacch
759	27	100.0	233	2	Q8C7U4_MOUSE	Q8c7u4 mus musculus	832	27	100.0	253	1	RS2_SCHPO	Q74892 schizosacch
760	27	100.0	233	2	Q7TFN9_RHCM6	Q7tfn9 rhesus cyto	833	27	100.0	253	1	RS3_CHLTE	Q8kah8 chlorobium
761	27	100.0	234	2	Q5VU80_HUMAN	Q5vu80 homo sapien	834	27	100.0	253	1	YB0XH_APLCA	P41824 aplysia cal

835	27	100.0	253	2	Q5B3B9_EMENI	Q5B3b9 aspergillus	908	27	100.0	266	2	Q23960_DUGJA	Q23960 dugesia jap
836	27	100.0	253	2	Q4N459_THERPA	Q4n459 theileria p	909	27	100.0	266	2	Q5QW66_ORYSA	Q5qwm66 oryza sativ
837	27	100.0	253	2	Q4UFV1_THERPA	Q4ufv1 theileria a	910	27	100.0	266	2	Q8BXC2_MOUSE	Q8bxc2 m mus muscu
838	27	100.0	253	2	Q6K4C2_ORYSA	Q6k4c2 oryza sativ	911	27	100.0	267	1	RS2_DROME	RS2 drosophila
839	27	100.0	253	2	Q6K7S2_ORYSA	Q6k7s2 oryza sativ	912	27	100.0	267	2	Q6XIL7_DROYA	Q6xil7 drosophila
840	27	100.0	253	2	Q8W4A3_ARATH	Q8w4a3 arabidopsis	913	27	100.0	267	2	Q5SN06_ORYSA	Q5sn06 oryza sativ
841	27	100.0	253	2	Q4NPB5_9DELT	Q4npb5 anaeromyxob	914	27	100.0	267	2	Q8LEU2_ARATH	Q8leu2 arabidopsis
842	27	100.0	254	1	THOC4_MOUSE	Q08583 mus musculus	915	27	100.0	267	2	Q9FY96_ARATH	Q9fy96 arabidopsis
843	27	100.0	254	2	Q5BGN2_EMENI	Q5bgn2 aspergillus	916	27	100.0	267	2	Q9LU62_ARATH	Q9lu62 arabidopsis
844	27	100.0	254	2	Q4WKD9_ASPFU	Q4wk9 aspergillus	917	27	100.0	267	2	Q4RXZ2_TETNG	Q4rxz2 tetraodon n
845	27	100.0	254	2	Q4HW64_GIBZE	Q4hw64 gibberella	918	27	100.0	268	2	Q9ZD17_RICPR	Q9zd17 rickettsia
846	27	100.0	254	2	Q5KJS7_CRYNE	Q5kjs7 cryptococcus	919	27	100.0	269	1	SFRS5_FAT	Q9d167 rattus norv
847	27	100.0	254	2	Q9H2F8_HUMAN	Q9h2f8 homo sapien	920	27	100.0	269	2	Q6C2T5_YARLI	Q6c2t5 yarrowia li
848	27	100.0	254	2	Q7R097_GIALA	Q7r097 giardia lam	921	27	100.0	269	2	Q4YT72_PLABE	Q4yt72 plasmodium
849	27	100.0	254	2	Q8W0W6_MAIZE	Q8w0w6 zea mays (m	922	27	100.0	269	2	Q651C3_ORYSA	Q651c3 oryza sativ
850	27	100.0	254	2	Q6A7H0_PROAC	Q6a7h0 propionibac	923	27	100.0	269	2	Q640L9_MOUSE	Q640l9 mus musculu
851	27	100.0	254	2	Q52155_NOCFA	Q52155 nocardia fa	924	27	100.0	270	1	SFRS5_MOUSE	Q5326 mus musculu
852	27	100.0	254	2	Q8K101_MOUSE	Q8k101 mus musculu	925	27	100.0	270	2	Q7QTKI_GIALA	Q7qtki giardia lam
853	27	100.0	255	2	Q5SR73_CRYNE	Q5sr73 cryptococcus	926	27	100.0	270	2	Q615B3_ORYSA	Q615b3 oryza sativ
854	27	100.0	255	2	Q94B27_ARATH	Q94b27 arabidopsis	927	27	100.0	270	2	Q6K2A7_ORYSA	Q6k2a7 oryza sativ
855	27	100.0	255	2	Q89F92_BRAJA	Q89f92 bradyrhizob	928	27	100.0	270	2	Q82R23_STRAW	Q82rj3 streptomyce
856	27	100.0	256	1	THOC4_HUMAN	Q86v81 homo sapien	929	27	100.0	270	2	Q9D8S5_MOUSE	Q9d8s5 mus musculu
857	27	100.0	256	2	Q8U673_AGR75	Q8u673 agrobacteri	930	27	100.0	271	2	Q7KZ74_HUMAN	Q7kz74 homo sapien
858	27	100.0	256	2	Q4RKZ5_TETNG	Q4rkz5 tetraodon n	931	27	100.0	271	2	O01644_DROGR	O01644 drosophila
859	27	100.0	256	2	Q6GLW1_XENLA	Q6glw1 xenopus lae	932	27	100.0	271	2	Q7QFZ1_ANOGA	Q7qfz1 anopheles g
860	27	100.0	256	2	Q58EA2_XENLA	Q58ea2 xenopus lae	933	27	100.0	271	2	Q4TPJ4_9SPHN	Q4tpj4 erythrobact
861	27	100.0	257	2	Q4IMP9_GIBZE	Q4imp9 gibberella	934	27	100.0	271	2	Q4K9P5_PSEF5	Q4k9p5 pseudomonas
862	27	100.0	257	2	Q5ZDJ8_ORYSA	Q5zdj8 oryza sativ	935	27	100.0	271	2	Q88FM3_PSRPK	Q88fm3 pseudomonas
863	27	100.0	257	2	Q6VYC2_ORYSA	Q6vyc2 oryza sativ	936	27	100.0	272	1	SFRS5_HUMAN	O13243 homo sapien
864	27	100.0	257	2	Q5YNK0_NOCFA	Q5ynk0 nocardia fa	937	27	100.0	272	1	Y366_HYCPN	P75415 mycoplasma
865	27	100.0	258	2	Q8W0W7_ORYSA	Q8w0w7 oryza sativ	938	27	100.0	272	2	Q9YEL6_AERPE	Q9y616 aeropyrum p
866	27	100.0	258	2	Q5B7R7_EMENI	Q5b7r7 aspergillus	939	27	100.0	272	2	Q8IL02_PLAF7	Q8il02 plasmodium
867	27	100.0	259	2	Q6C5W9_YARLI	Q6c5w9 yarrowia li	940	27	100.0	272	2	Q6YX24_ORYSA	Q6yx24 oryza sativ
868	27	100.0	259	2	Q4WA18_ASPFU	Q4wa18 aspergillus	941	27	100.0	272	2	Q8H7C6_ARATH	Q8h7c6 arabidopsis
869	27	100.0	259	2	Q5VWF8_ORYSA	Q5vmf8 oryza sativ	942	27	100.0	272	2	Q93H15_STRAW	Q93h15 streptomyce
870	27	100.0	259	2	Q7OJPE_WHEAT	Q7ojpe triticum ae	943	27	100.0	273	2	Q5MIQ3_ABDAL	Q5miq9 aedes albop
871	27	100.0	259	2	Q7NV98_CHRVO	Q7nv98 chromobacte	944	27	100.0	273	2	Q4KTE1_SUBDO	Q4kte1 suberites d
872	27	100.0	260	2	Q96BS4_HUMAN	Q96bs4 homo sapien	945	27	100.0	273	2	Q65X75_ORYSA	Q65x75 oryza sativ
873	27	100.0	260	2	Q9BSF4_HUMAN	Q9bsf4 homo sapien	946	27	100.0	274	2	Q6Z4L2_ORYSA	Q6z4l2 oryza sativ
874	27	100.0	260	2	Q86ER5_SCHJA	Q86er5 schistosoma	947	27	100.0	275	2	Q4N7F1_THEPA	Q4n7f1 theileria p
875	27	100.0	261	2	Q27938_METTH	Q27938 methanobact	948	27	100.0	275	2	Q6L3I8_SOLDE	Q6l3i8 solanum dem
876	27	100.0	261	2	Q5UAP3_BOMMO	Q5uap3 bombyx mori	949	27	100.0	275	2	Q8CK25_STRCO	Q8ck25 streptomyce
877	27	100.0	261	2	Q6EV29_PAPDA	Q6ev29 papilio dar	950	27	100.0	275	2	Q8FRM1_COREF	Q8frm1 corynebacte
878	27	100.0	261	2	Q81R65_SPOFR	Q81r65 spodoptera	951	27	100.0	275	2	Q4T647_TETNG	Q4t647 tetraodon n
879	27	100.0	261	2	Q6MW91_ORYSA	Q6mw91 oryza sativ	952	27	100.0	276	2	Q51NZ6_MAGGR	Q51nz6 magnaporthe
880	27	100.0	261	2	Q5QLV6_ORYSA	Q5qlv6 oryza sativ	953	27	100.0	276	2	Q5JV08_HUMAN	Q5jv08 homo sapien
881	27	100.0	261	2	Q8U8B6_AGR75	Q8u8b6 agrobacteri	954	27	100.0	276	2	Q9SCM3_ARATH	Q9scm3 arabidopsis
882	27	100.0	261	2	Q56A61_XENLA	Q56a61 xenopus lae	955	27	100.0	277	2	Q4WPA4_ASPFU	Q4wpa4 aspergillus
883	27	100.0	262	2	Q7RA04_PLAYO	Q7ra04 plasmodium	956	27	100.0	277	2	Q8LI48_ORYSA	Q8li48 oryza sativ
884	27	100.0	262	2	Q652Z0_ORYSA	Q652z0 oryza sativ	957	27	100.0	277	2	Q4SVK9_TETNG	Q4svk9 tetraodon n
885	27	100.0	262	2	Q6YY32_ORYSA	Q6yy32 oryza sativ	958	27	100.0	278	1	RS2_URECA	P49154 urechis cau
886	27	100.0	262	2	Q7XUT4_ORYSA	Q7xut4 oryza sativ	959	27	100.0	278	2	Q9H7J7_HUMAN	Q9h7j7 homo sapien
887	27	100.0	262	2	Q7NNM0_GLOVI	Q7nnm0 gloebacter	960	27	100.0	278	2	Q12845_HUMAN	Q12845 homo sapien
888	27	100.0	263	2	Q8X0W3_NEUCR	Q8x0w3 neurospora	961	27	100.0	279	1	MAGG1_MOUSE	Q9cpr8 mus musculu
889	27	100.0	264	1	Y1554_AQUAE	Q67506 aquifex ae	962	27	100.0	279	2	Q9FW02_ARATH	Q9fw02 arabidopsis
890	27	100.0	264	2	Q875D8_PODAN	Q875d8 podospora a	963	27	100.0	279	2	Q4IWO7_AZOVI	Q4iwo7 azotobacter
891	27	100.0	264	2	Q4Q5P0_LEIMA	Q4q5p0 leishmania	964	27	100.0	280	2	Q5VU78_HUMAN	Q5vu78 homo sapien
892	27	100.0	264	2	Q4IZZ2_AZOVI	Q4izz2 azotobacter	965	27	100.0	280	2	Q6NWC3_BRARE	Q6nwc3 brachydanio
893	27	100.0	264	2	Q63U70_BURPS	Q63u70 burkholderi	966	27	100.0	280	2	Q76DT5_ONCMA	Q76dt5 oncorhynch
894	27	100.0	264	2	Q6ZKD6_BURMA	Q6zkd6 burkholderi	967	27	100.0	280	2	Q76DRT_ONCMA	Q76drt oncorhynch
895	27	100.0	265	1	RS2_LEIAM	Q43992 leishmania	968	27	100.0	281	2	Q7S2R3_NEUCR	Q7s2r3 neurospora
896	27	100.0	265	2	Q5EMW4_MAGGR	Q5emw4 magnaporthe	969	27	100.0	281	2	Q6Z633_ORYSA	Q6z633 oryza sativ
897	27	100.0	265	2	Q7S503_NEUCR	Q7s503 neurospora	970	27	100.0	281	2	Q87598_STRCO	Q87598 streptomyce
898	27	100.0	265	2	Q52514_MAGGR	Q52514 magnaporthe	971	27	100.0	281	2	Q7AKQ5_STRCO	Q7akq5 streptomyce
899	27	100.0	265	2	Q4ILR0_GIBZE	Q4ilr0 gibberella	972	27	100.0	281	2	Q68ER8_XENTR	Q68er8 xenopus tro
900	27	100.0	265	2	Q5VU79_HUMAN	Q5vu79 homo sapien	973	27	100.0	281	2	Q7T0R9_XENLA	Q7t0r9 xenopus lae
901	27	100.0	265	2	Q71LW5_LEIMA	Q71lw5 leishmania	974	27	100.0	281	2	Q4RJH8_TETNG	Q4rjh8 tetraodon n
902	27	100.0	265	2	Q4QDL5_LEIMA	Q4qdl5 leishmania	975	27	100.0	282	2	Q7SCA9_NEUCR	Q7sca9 neurospora
903	27	100.0	265	2	Q942B5_ORYSA	Q942b5 oryza sativ	976	27	100.0	282	2	Q9SM06_ARATH	Q9sm06 arabidopsis
904	27	100.0	265	2	Q91298_PSEAE	Q91298 pseudomonas	977	27	100.0	282	2	Q9C7U4_ARATH	Q9c7u4 arabidopsis
905	27	100.0	265	2	Q6DDA7_XENTR	Q6dda7 xenopus tro	978	27	100.0	282	2	Q881Z2_PSESM	Q881z2 pseudomonas
906	27	100.0	265	2	Q4S9L1_TETNG	Q4s9l1 tetraodon n	979	27	100.0	283	2	O07050_AERHY	O07050 aeromonas h
907	27	100.0	266	1	SLU7_ASHGO	Q74zn9 ashbya goss	980	27	100.0	283	2	Q537D4_MOUSE	Q537d4 mus musculu

RN	[1]									
RP	NUCLEOTIDE SEQUENCE.									
RA	Han Z.;									
RL	Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.									
DR	EMBL; AY812248; AAX28137.1; -; mRNA.									
KW	Hypothetical protein.									
SQ	SEQUENCE 30 AA; 3534 MW; 42D8142D91F9A29F CRC64;									
Query Match 100.0%; Score 27; DB 2; Length 30;										
Best Local Similarity 100.0%; Pred. No. 1.4e+02;										
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;										
QY	1 RGRGR 5									
Db										
RESULT 3										
ID	Q7TEPH0_MOUSE PRELIMINARY; PRT; 35 AA.									
AC	Q7TEPH0;									
DT	01-OCT-2003 (Tremblrel. 25, Created)									
DT	01-OCT-2003 (Tremblrel. 25, Last sequence update)									
DE	FGF2 (Fragment).									
GN	Name=FGF2;									
OS	Mus musculus (Mouse).									
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;									
OC	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;									
OC	Muridae; Murinae; Mus.									
OX	NCBI_TaxID=10090;									
RN	[1]									
RP	NUCLEOTIDE SEQUENCE.									
RC	STRAIN=129/SVJ;									
RX	PubMed=14618271; DOI=10.1007/s00018-003-3258-6;									
RA	Folletti A., Vuadens F., Beermann F.;									
RT	"Nuclear localization of mouse fibroblast growth factor 2 requires N-terminal and C-terminal sequences.";									
RL	Cell. Mol. Life Sci. 60:2254-2265(2003).									
DR	EMBL; AY324448; AAP92384.1; -; Genomic_DNA.									
DR	GO; GO:0005615; C:extracellular space; IDA.									
DR	GO; GO:0005515; F:protein binding; IPI.									
DR	GO; GO:001525; P:angiogenesis; IDA.									
DR	GO; GO:0010001; P:glial cell differentiation; IMP.									
DR	GO; GO:0001759; P:induction of an organ; IDA.									
DR	GO; GO:0030324; P:lung development; IDA.									
DR	GO; GO:0045597; P:positive regulation of cell differentiation; IDA.									
DR	GO; GO:0046668; P:regulation of retinal programmed cell death; IDA.									
FT	NON_TER 35 35									
SQ	SEQUENCE 35 AA; 3451 MW; 08F3AF21ABFC7250 CRC64;									
Query Match 100.0%; Score 27; DB 2; Length 35;										
Best Local Similarity 100.0%; Pred. No. 1.7e+02;										
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;										
QY	1 RGRGR 5									
Db										
RESULT 4										
ID	P78343_HUMAN PRELIMINARY; PRT; 36 AA.									
AC	P78343;									
DT	01-MAY-1997 (Tremblrel. 03, Created)									
DT	01-MAY-1997 (Tremblrel. 03, Last sequence update)									
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)									
DE	Factor XII (Fragment).									
OS	Homo sapiens (Human).									
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;									
OC	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae;									
OC	Homo.									
OX	NCBI_TaxID=9606;									

981	27	100.0	284	2	Q7EYN6	oryza sativ
982	27	100.0	284	2	Q8L861	arabidopsis
983	27	100.0	284	2	Q8L8Y0	arabidopsis
984	27	100.0	284	2	Q93VB8	arabidopsis
985	27	100.0	284	2	Q9SM07	arabidopsis
986	27	100.0	284	2	Q4TP26	erythrobact
987	27	100.0	284	2	Q4TOR0	tetraodon n
988	27	100.0	284	2	Q6NUT4	brachydanio
989	27	100.0	285	2	Q6H8Z4	anopheles g
990	27	100.0	285	2	Q56FF2	lysiphlebus
991	27	100.0	285	2	Q7W1M3	borderella
992	27	100.0	285	2	Q7WPL4	borderella
993	27	100.0	286	1	R82	bov
994	27	100.0	286	2	Q59ZV3	canal
995	27	100.0	286	2	Q5TR65	anopheles g
996	27	100.0	286	2	Q9U2U9	caenorhabdi
997	27	100.0	286	2	Q60YC3	caenorhabdi
998	27	100.0	286	2	O82021	medicago sa
999	27	100.0	286	2	Q7EY69	oryza sativ
1000	27	100.0	287	2	Q5B7S1	EMENI
ALIGNMENTS						
RESULT 1						
ID	Q700R5_PIG PRELIMINARY; PRT; 29 AA.					
AC	Q700R5;					
DT	05-JUL-2004 (Tremblrel. 27, Created)					
DT	05-JUL-2004 (Tremblrel. 27, Last sequence update)					
DE	High mobility group protein 1 (Fragment).					
GN	Name=HMG1;					
OS	Sus scrofa (Pig).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;					
OC	Sus.					
OX	NCBI_TaxID=9823;					
RN	[1]					
RP	NUCLEOTIDE SEQUENCE.					
RA	Barbosa A., Demeure O., Urien C., Milan D., Chardon P., Renard C.;					
RT	"A physical map of large segments of pig chromosome 7ql1-q14;					
RT	comparative analysis with human chromosome 6p21.";					
RL	Mamm. Genome 15:982-995(2004).					
DR	EMBL; AJ629180; CAF32801.1; -; Genomic_DNA.					
DR	InterPro: IPR000637; A+T hook.					
DR	PROSITE; PS00354; HMG1_Y; 1.					
FT	NON_TER 1 1					
FT	NON_TER 29 29					
SQ	SEQUENCE 29 AA; 3131 MW; 6BC19DD6744755E7 CRC64;					
Query Match 100.0%; Score 27; DB 2; Length 29;						
Best Local Similarity 100.0%; Pred. No. 1.4e+02;						
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;						
QY	1 RGRGR 5					
Db						
RESULT 2						
ID	Q5BWT8_SCHJA PRELIMINARY; PRT; 30 AA.					
AC	Q5BWT8;					
DT	10-MAY-2005 (Tremblrel. 30, Created)					
DT	10-MAY-2005 (Tremblrel. 30, Last sequence update)					
DT	10-MAY-2005 (Tremblrel. 30, Last annotation update)					
DE	Hypothetical protein.					
OS	Schistosoma japonicum (Blood fluke).					
OC	Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;					
OC	Schistosomatidae; Schistosomatidae; Schistosoma.					
OX	NCBI_TaxID=6182;					



```

RN NUCLEOTIDE SEQUENCE.
RC TISSUE=Blood;
RX MEDLINE=88007593; PubMed=2888762;
RA Cool D.E., McGillivray R.T.A.;
RT "Characterization of the human blood coagulation factor XII gene.
RT Intron/exon gene organization and analysis of the 5'-flanking
RT region.";
RL J. Biol. Chem. 262:13662-13673(1987).
RN NUCLEOTIDE SEQUENCE.
RC TISSUE=Blood;
RX MEDLINE=96133302; PubMed=8528215;
RA Schloesser M., Hofferbert S., Bartz U., Lutze G., Lammle B., Engel W.;
RT "The novel acceptor splice site mutation 11396(G->A) in the factor
RT XII gene causes a truncated transcript in cross-reacting material
RT negative patients.";
RL Hum. Mol. Genet. 4:1235-1237(1995).
RN NUCLEOTIDE SEQUENCE.
RC TISSUE=Blood;
RX MEDLINE=96235284; PubMed=8641707; DOI=10.1007/s004390050147;
RA Hofferbert S., Muller J., Kosterling H., von Ohlen W.D., Schloesser M.;
RT "A novel 5'-upstream mutation in the factor XII gene is associated
RT with a TaqI restriction site in an Alu repeat in factor XII-deficient
RT patients.";
RL Hum. Genet. 97:838-841(1996).
RN NUCLEOTIDE SEQUENCE.
RC TISSUE=Blood;
RA Hinney B., Koesterling H., Laemmle B., Pindur G., Thies K.,
RA Schloesser M., Zeerleder C., Lutze G., Halbmayer W.M., Hofferbert S.,
RA Koehler M., Engel W.A.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RT EMBL; U71278; AAB51207.1; -; Genomic_DNA.
FT NON_TER 1
SQ SEQUENCE 36 AA; 3577 MW; 5D133B35BC04304F CRC64;

Query Match 100.0%; Score 27; DB 2; Length 36;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
Db 28 RGRGR 32

RESULT 5
Q9J154 RAT PRELIMINARY; PRT; 37 AA.
AC Q9J154;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE High mobility group protein HMGI-C (Fragment).
GN Name=Hmgi-C;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=Sprague-Dawley;
RA Wen X., Lin H.H., Deng H.-T., Hajjyan K., Ann D.K.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF261719; AAF91385.1; -; Genomic_DNA.
DR GO; GO:000785; C:chromatin; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0007001; P:chromosome organization and biogenesis (sen. ...; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000637; A+T_hook.

```

```

DR InterPro; IPR000116; Highmoblty_IY.
DR PRINTS; PRO0929; ATHOOK.
DR ProDom; PD005593; Highmoblty_IY; 1.
FT PROSITE; PS00354; HMGI_Y; 1.
FT NON_TER 37
SQ SEQUENCE 37 AA; 3828 MW; 7E23A16F3EB01AC2 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 37;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
Db 27 RGRGR 31

RESULT 6
Q56X91 ARATH PRELIMINARY; PRT; 45 AA.
AC Q56X91;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DE Hypothetical protein Atig76010 (Fragment).
GN Name=Atig76010;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC euroside II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN NUCLEOTIDE SEQUENCE.
RP Totoki Y., Seki M., Ishida J., Nakajima M., Enju A., Kamiya A.,
RA Narusaka M., Shin-i T., Nakagawa M., Sakamoto N., Oishi K., Kohara Y.,
RA Kobayashi M., Toyoda A., Sakaki Y., Sakurai T., Iida K., Akiyama K.,
RA Satou M., Toyoda T., Konagaya A., Carninci P., Kawai J.,
RA Hayashizaki Y., Shinozaki K.;
RT "Large-scale analysis of RIKEN Arabidopsis full-length (RAFL) cDNAs.";
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK221785; BAD93912.1; -; mRNA.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 45 AA; 4385 MW; 656D6DDAD96F751F CRC64;

Query Match 100.0%; Score 27; DB 2; Length 45;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
Db 9 RGRGR 13

RESULT 7
Q6L736_9ACTO PRELIMINARY; PRT; 51 AA.
AC Q6L736;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein.
OS Streptomyces kanamyceticus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1967;
RN NUCLEOTIDE SEQUENCE.
RP Yanai K., Murakami T.;
RA "The kanamycin biosynthetic gene cluster from Streptomyces
RT kanamyceticus.";
RL J. Antibiot. 57:351-354(2004).
DR EMBL; AB164642; BAD20761.1; -; Genomic_DNA.
KW Hypothetical protein.

```

```
SQ SEQUENCE 51 AA; 5200 MW; 361B434049592A31 CRC64;
Query Match 100.0%; Score 27; DB 2; Length 51;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
    |||||
Db 29 RGRGR 33

RESULT 8
Q9S2H2_STRCO PRELIMINARY; PRT; 55 AA.
AC Q9S2H2_
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative membrane protein.
GN OrderedLocNames=SCO2909; ORFNames=SC19A.09;
OS Streptomyces coelicolor
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953; DOI=10.1038/417141a;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieffer T., Larkie L., Murphy L.D., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K.M., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J.R., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL939114; CAB50990.1; -; Genomic_DNA.
DR FIC; T36132; T36132.
KW Complete proteome.
SQ SEQUENCE 55 AA; 5971 MW; C5C2BECDB0A253F9 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 55;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
    |||||
Db 32 RGRGR 36

RESULT 9
Q6YYD0_ORYSA PRELIMINARY; PRT; 56 AA.
AC Q6YYD0_
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Hypothetical protein P0680F05.31 (Hypothetical protein
DE P0509D04.38)
GN Name=P0680F05.31; Synonyms=P0509D04.38;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Katayose Y.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBSJ databases.

SQ SEQUENCE 51 AA; 5200 MW; 361B434049592A31 CRC64;
Query Match 100.0%; Score 27; DB 2; Length 51;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
    |||||
Db 29 RGRGR 33

RESULT 8
Q9S2H2_STRCO PRELIMINARY; PRT; 55 AA.
AC Q9S2H2_
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative membrane protein.
GN OrderedLocNames=SCO2909; ORFNames=SC19A.09;
OS Streptomyces coelicolor
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953; DOI=10.1038/417141a;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieffer T., Larkie L., Murphy L.D., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K.M., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J.R., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL939114; CAB50990.1; -; Genomic_DNA.
DR FIC; T36132; T36132.
KW Complete proteome.
SQ SEQUENCE 55 AA; 5971 MW; C5C2BECDB0A253F9 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 55;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
    |||||
Db 32 RGRGR 36

RESULT 9
Q6YYD0_ORYSA PRELIMINARY; PRT; 56 AA.
AC Q6YYD0_
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Hypothetical protein P0680F05.31 (Hypothetical protein
DE P0509D04.38)
GN Name=P0680F05.31; Synonyms=P0509D04.38;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBSJ databases.

SQ SEQUENCE 56 AA; 6234 MW; D7AFA17BC805056F CRC64;
Query Match 100.0%; Score 27; DB 2; Length 56;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
    |||||
Db 26 RGRGR 30

RESULT 10
Q69128_9GAMA PRELIMINARY; PRT; 58 AA.
AC Q69128_
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Epstein-Barr virus (strain FF41) nuclear antigen gene (EBNA 1),
DE partial cds. (Fragment).
OS Human herpesvirus 4 (Epstein-Barr virus).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10376;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=86200373; PubMed=3009849;
RA Polivino-Bodnar M., Shedd D., Miller G.;
RA "Deletion mutants that affect expression of Epstein-Barr virus nuclear
RT antigen in COS-1 cells after gene transfer with simian virus 40
RT vectors containing portions of the BamHI K fragment.";
RL J. Virol. 58:324-330(1986).
DR EMBL; M13180; AAA45883.1; -; Genomic_DNA.
FT NON TER 58
SQ SEQUENCE 58 AA; 5440 MW; 42367997A0353A11 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 58;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
    |||||
Db 41 RGRGR 45

RESULT 11
Q6ZFA3_ORYSA PRELIMINARY; PRT; 59 AA.
AC Q6ZFA3_
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein OJ1224_G08.15.
GN Name=OJ1224_G08.15;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 8, BAC
RT clone:OJ1224_G08.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AP004256; BAD01684.1; -; Genomic_DNA.
DR Gramene; Q6ZFA3; -.
KW Hypothetical protein.
SQ SEQUENCE 59 AA; 6324 MW; 2C84938755983AA4 CRC64;
```

Query Match 100.0%; Score 27; DB 2; Length 59;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
 DB 34 RGRGR 38

## RESULT 12

O64001\_9MURI  
 ID O64001\_9MURI PRELIMINARY; PRT; 59 AA.  
 AC O64001;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-WAR-2004 (TREMBLrel. 26, Last annotation update)  
 DE SMD homolog protein (Fragment).  
 GN Name-SMD homolog;  
 OS Mus sp.  
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10095;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=94257110; PubMed=7515243; DOI=10.1006/jaut.1994.1009;  
 RA Rivkin E., Vella M.J., Lahita R.G.;  
 RT "A heterogeneous immune response to an SMD-like epitope by SLE patients."  
 RT J. Autoimmun. 7:119-132(1994).  
 RL EMBL; S71494; AAB30912.1; -; mRNA.  
 DR InterPro; IPR000637; A+T\_hook.  
 DR PRINTS; PR00929; ATHOOK.  
 FT NON\_TER 59  
 FT SEQUENCE 59 AA; 6547 MW; C95BBEBE98519054A CRC64;

Query Match 100.0%; Score 27; DB 2; Length 59;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
 DB 17 RGRGR 21

## RESULT 13

O6QX44\_9VIRU  
 ID O6QX44\_9VIRU PRELIMINARY; PRT; 59 AA.  
 AC O6QX44;  
 DT 05-JUL-2004 (TREMBLrel. 27, Created)  
 DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)  
 DE Capsid protein (Fragment).  
 GN Name-ORF2;  
 OS Human astrovirus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Astroviridae;  
 OC Mamastrovirus.  
 OX NCBI\_TaxID=12702;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=AstCS2;  
 RA Grimm A., Cashdollar J., Williams F., Fout G.S.;  
 RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY522552; AAS49170.1; -; Genomic\_RNA.  
 DR InterPro; IPR004337; Astro\_capsid.  
 DR Pfam; PF03115; Astro\_capsid; 1.  
 FT NON\_TER 1  
 FT NON\_TER 59  
 FT SEQUENCE 59 AA; 6985 MW; EA167D0C780BE966 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 59;  
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RGRGR 5  
 DB 17 RGRGR 21

## RESULT 14

Q80VMO\_MOUSE  
 ID Q80VMO\_MOUSE PRELIMINARY; PRT; 61 AA.  
 AC Q80VMO;  
 DT 01-JUN-2003 (TREMBLrel. 24, Created)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Similar to thyroid hormone receptor-associated protein, 150 kDa subunit.  
 DE Mus musculus (Mouse).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=olfactory epithelium;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=olfactory epithelium;  
 RA Strausberg R.;  
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC048391; AAH48391.1; -; mRNA.  
 DR Ensembl; ENSMUSG0000052460; Mus musculus.  
 DR GO; GO:0004872; F:receptor activity; IEA.  
 KW Receptor.  
 SQ SEQUENCE 61 AA; 6875 MW; DA8EF77EBED40411 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 61;  
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
 DB 3 RGRGR 7

## RESULT 15

Q7QWGB\_GIALA  
 ID Q7QWGB\_GIALA PRELIMINARY; PRT; 62 AA.  
 AC Q7QWGB;  
 DT 01-WAR-2004 (TREMBLrel. 26, Created)  
 DT 01-WAR-2004 (TREMBLrel. 26, Last sequence update)  
 DT 01-WAR-2004 (TREMBLrel. 26, Last annotation update)  
 DE GLP\_336\_614\_802.

```

OS Giardia lamblia ATCC 50803.
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiae; Giardia.
OX NCBI_TaxID=184922;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=WB C6;
RA Morrison H.G., McArthur A.G., Adam R.D., Aley S.B., Gillin F.D.,
RA Olsen G.J., Sogin M.L.;
RT "Draft sequence of the Giardia lamblia genome.";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
CC "-!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AACB01000071; EAA39354.1; -; Genomic DNA.
SQ SEQUENCE 62 AA; 6980 MW; 5CFBB1A193BB77DB CRC64;

Query Match 100.0%; Score 27; DB 2; Length 62;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
DB 31 RGRGR 35

RESULT 16
Q626R0_ORYSA
ID Q626R0_ORYSA PRELIMINARY; PRT; 65 AA.
AC Q626R0;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DE Hypothetical protein P0705A04.30.
GN Name=P0705A04.30;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa niponbare (GA3) genomic DNA, chromosome 2, PAC
RT clone:P0705A04.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP004891; BAD15985.1; -; Genomic DNA.
DR Gramene; Q626R0; -;
KW Hypothetical protein.
SQ SEQUENCE 65 AA; 6969 MW; 2A670FC9D048D0EF CRC64;

Query Match 100.0%; Score 27; DB 2; Length 65;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
DB 47 RGRGR 51

RESULT 17
Q919P6_9BACU
ID Q919P6_9BACU PRELIMINARY; PRT; 65 AA.
AC Q919P6;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE CUN023 putative p6.9 SR repeat DNA binding protein, similar to AcnNPV
DE ORF100.
GN Name=CUN023;
OS Culex nigripalpus baculovirus.
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae.
OX NCBI_TaxID=130556;
RN [1]

```

```

RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21488685; PubMed=11602755;
RX DOI=10.1128/JVI.75.22.11157-11165.2001;
RA Afonso C.L., Tulman E.R., Lu Z., Balinsky C.A., Moser B.A.,
RA Becnel J.J., Rock D.L., Kutish G.F.;
RT "Genome sequence of a baculovirus pathogenic for Culex nigripalpus.";
RL J. Virol. 75:11157-11165(2001).
DR EMBL; AF403738; AAK94101.1; -; Genomic DNA.
SQ SEQUENCE 65 AA; 7488 MW; 4F4521FC766CB2EB CRC64;

Query Match 100.0%; Score 27; DB 2; Length 65;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
DB 59 RGRGR 63

RESULT 18
Q7NZG4_CHRVO
ID Q7NZG4_CHRVO PRELIMINARY; PRT; 67 AA.
AC Q7NZG4;
DT 01-MAR-2004 (TREMBlrel. 26, Created)
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Probable tisl421-transposase orfa protein.
GN OrderedLocustNames=CV0958;
OS Chromobacterium violaceum.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Chromobacterium.
OX NCBI_TaxID=536;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 12472 / DSM 30191;
RX MEDLINE=22882880; PubMed=14500782; DOI=10.1073/pnas.1832124100;
RA Vasconcelos A.T.F., de Almeida D.F., Hungria M., Guimaraes C.T.,
RA Antonio R.V., Almeida F.C., de Almeida L.G.P., de Almeida R.,
RA Alves-Gomes J.A., Andrade E.M., Araripe J., de Araujo M.F.F.,
RA Astolfi-Filho S., Azevedo V., Baptista A.J., Bataus L.A.M.,
RA Batista J.S., Beio A., van den Berg C., Bogo M., Bonatto S.,
RA Bordignon J., Brigido M.M., Brito C.A., Brocchi M., Buriti H.A.,
RA Canargo A.A., Cardoso D.D.P., Carneiro N.P., Carraro D.M.,
RA Carvalho C.M.B., Cascardo J.C.M., Cavada B.S., Chueire L.M.O.,
RA Creczynski-Pasa T.B., Cunha-Junior N.C., Fagundes N., Falcao C.L.,
RA Fantinatti F., Farias I.P., Felipe M.S.S., Fertari L.P., Ferro J.A.,
RA Ferro M.I.T., Franco G.R., Freitas N.S.A., Furian L.R.,
RA Gazzinelli R.T., Gomes E.A., Goncalves P.R., Grangeiro T.B.,
RA Grattapaglia D., Grisard E.C., Hanna E.S., Jardim S.N., Laurino J.,
RA Leoi L.C.T., Lima L.F.A., Loureiro M.F., Lyra M.C.C.P.,
RA Madeira H.M.F., Manfio G.P., Maranhao A.Q., Martins W.S.,
RA di Mauro S.M.Z., de Medeiros S.R.B., Meissner R.V., Moreira M.A.M.,
RA Nascimento F.F., Nicolas M.F., Oliveira J.G., Oliveira S.C.,
RA Paixao R.F.C., Parente J.A., Pedrosa F.O., Pena S.D.J., Pereira J.O.,
RA Pereira M., Pinto L.S.R.C., Pinto L.S., Porto J.I.R., Potrich D.P.,
RA Ramalho-Neto C.E., Reis A.M.M., Rigo L.U., Rondinelli E.,
RA Santos E.B.P., Santos P.R., Schneider M.P.C., Seunanez H.N.,
RA Silva A.M.R., da Silva A.L.C., Silva D.W., Silva R., Simoes I.C.,
RA Simon D., Soares C.M.A., Soares R.B.A., Souza E.M., Souza K.R.L.,
RA Souza R.C., Steffens M.B.R., Steindel M., Teixeira S.R., Urmenyi T.,
RA Vettore A., Wassen R., Zaha A., Simpson A.J.G.;
RT "The complete genome sequence of Chromobacterium violaceum reveals
RT remarkable and exploitable bacterial adaptability.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:11660-11665(2003).
DR EMBL; AE016913; AAQ58632.1; -; Genomic DNA.
KW Complete proteome.
SQ SEQUENCE 67 AA; 7697 MW; 6FA1C11710A093D7 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 67;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5

```

```
DB 25 RGRGR 29
|||||
RESULT 19
HSPI_TACAC STANDARD; PRT; 68 AA.
AC P3511;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Sperm protamine P1.
GN Names=PRM1; Synonyms=PRM-1;
OS Tachyglossus aculeatus aculeatus (Australian echidna).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Monotremata; Tachyglossidae; Tachyglossus.
OX NCBI_TaxID=49271;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=94094837; PubMed=8269934;
RA Retief J.D., Winkfein R.J., Dixon G.H.;
RT "Evolution of the monotremes. The sequences of the protamine P1 genes
of platypus and echidna.";
RL Eur. J. Biochem. 218:457-461(1993).
CC -!- FUNCTION: Protamines substitute for histones in the chromatin of
sperm during the haploid phase of spermatogenesis. They compact
sperm DNA into a highly condensed, stable and inactive complex.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: Testis.
CC -!- SIMILARITY: Belongs to the protamine P1 family.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use as long as its content is in no way modified and this statement is not
removed.
CC
CC EMBL; Z26848; CAA81444.1; -; Genomic_DNA.
DR PIR; S39424; S39424.
DR InterPro; IPR000221; Protamine_P1.
DR Pfam; PF00260; Protamine_P1; 1.
DR PROSITE; PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; Developmental protein; Differentiation;
DNA condensation; DNA-binding; Nuclear protein; Nucleosome core;
Spermatogenesis; Testis.
FT INIT_MET 0 0 By similarity.
SQ SEQUENCE 68 AA; 8673 MW; 41AEBF9BB212F7B3 CRC64;

Query Match 100.0%; Score 27; DB 1; Length 68;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
DB 40 RGRGR 44
|||||

RESULT 20
Q8WRX4 DROME
ID Q8WRX4 DROME PRELIMINARY; PRT; 68 AA.
AC Q8WRX4;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cytoplasmic tyrosine kinase (fragment).
GN Names=Fps85D; Synonyms=fer; ORFNames=CG8874;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP NUCLEOTIDE SEQUENCE.

RA Hill K.K., Bishop J.M.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF392049; AAL60138.1; -; Genomic_DNA.
DR FlyBase; FBgn0000723; CG8874.
DR FlyBase; FBgn0000723; Fps85D.
DR GO; GO:0005886; C:plasma membrane; IDA.
KW Kinase.
FT NON_TER 1 1
FT NON_TER 68 68
SQ SEQUENCE 68 AA; 7565 MW; C383F584EB7D18B6 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 68;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
DB 28 RGRGR 32
|||||

RESULT 21
Q5Z559 ORYSA
ID Q5Z559 ORYSA PRELIMINARY; PRT; 68 AA.
AC Q5Z559;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein OSJNBa0085C03.13.
GN Names=OSJNBa0085C03.13;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Katayose Y.;
RL "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, BAC
clone:OSJNBa0085C03.";
RT Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF005773; BAD62135.1; -; Genomic_DNA.
DR InterPro; IPR005829; Sug.transporter.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 68 AA; 7157 MW; B79EF60B678E5916 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 68;
Best Local Similarity 100.0%; Pred. No. 3.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
DB 62 RGRGR 66
|||||

RESULT 22
Q6ER06 ORYSA
ID Q6ER06 ORYSA PRELIMINARY; PRT; 68 AA.
AC Q6ER06;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Hypothetical protein P0701F11.32 (Hypothetical protein
P0668D04.13).
DE P0668D04.13).
GN Names=P0701F11.32; Synonyms=P0668D04.13;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Katayose Y.;
```

RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 RN NUCLEOTIDE SEQUENCE.  
 RA Sasaki T., Matsumoto T., Katayose Y.;  
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 9, PAC  
 clone: P0668D04.";  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP005429; BAD28664.1; -; Genomic DNA.  
 DR EMBL; AP005426; BAD33490.1; -; Genomic DNA.  
 DR Gramene; QSER06; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 68 AA; 7469 MW; 39921B0B4C2E0062 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 68;

Best Local Similarity 100.0%; Pred. No. 3.3e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5

Db 17 RGRGR 21

# RESULT 23

Q6Z0B8 ORYSA PRELIMINARY; PRT; 68 AA.  
 AC Q6Z0B8;  
 DT 05-JUL-2004 (TREMELrel. 27, Created)  
 DT 05-JUL-2004 (TREMELrel. 27, Last sequence update)  
 DT 01-FEB-2005 (TREMELrel. 29, Last annotation update)  
 DE Hypothetical protein OSJNB0003H03.1 (Hypothetical protein OSJNB0049G15.20).  
 GN Name=OSJNB003H03.1; Synonyms=OSJNB0049G15.20;  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzeae; Oryza.  
 OX NCBI\_TaxID=39947;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Sasaki T., Matsumoto T., Katayose Y.;  
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 8, BAC  
 clone: OSJNB0003H03.";  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Sasaki T., Matsumoto T., Katayose Y.;  
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 8, BAC  
 clone: OSJNB0049G15.";  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP005495; BAD05666.1; -; Genomic DNA.  
 DR EMBL; AP005064; BAD05494.1; -; Genomic DNA.  
 DR Gramene; Q6Z0B8; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 68 AA; 7575 MW; 1D099804419ED5CE CRC64;

Query Match 100.0%; Score 27; DB 2; Length 68;

Best Local Similarity 100.0%; Pred. No. 3.3e+02; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5

Db 52 RGRGR 56

# RESULT 24

Q6IK40 DROME PRELIMINARY; PRT; 74 AA.  
 AC Q6IK40;  
 DT 05-JUL-2004 (TREMELrel. 27, Created)  
 DT 05-JUL-2004 (TREMELrel. 27, Last sequence update)  
 DE Hypothetical protein.  
 GN ORFNames=HDC13425;

OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX PubMed14709175; DOI=10.1186/gb-2003-5-1-r3;  
 RA Hild M., Beckmann B., Haas S.A., Koch B., Solovyev V., Busold C.,  
 RA Pellenberg K., Boutros M., Vingron M., Sauer F., Hoheisel J.D.,  
 RA Paro R.;  
 RT "An integrated gene annotation and transcriptional profiling approach  
 towards the full gene content of the Drosophila genome.";  
 RL Genome Biol. 5:RESEARCH0003.1-RESEARCH0003.17(2003).  
 CC -!- MISCELLANEOUS: The sequence shown here is derived from an  
 EMBL/GenBank/DBJ third party annotation (TPA) entry.  
 DR EMBL; BK002526; DAA04032.1; -; Genomic DNA.  
 SQ SEQUENCE 74 AA; 8260 MW; F476D3C6B10BC291 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 74;

Best Local Similarity 100.0%; Pred. No. 3.6e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5

Db 33 RGRGR 37

# RESULT 25

Q5VQE4 ORYSA PRELIMINARY; PRT; 76 AA.  
 ID Q5VQE4;  
 AC Q5VQE4;  
 DT 01-FEB-2005 (TREMELrel. 29, Created)  
 DT 01-FEB-2005 (TREMELrel. 29, Last sequence update)  
 DT 01-FEB-2005 (TREMELrel. 29, Last annotation update)  
 DE Hypothetical protein OJ1316\_H05.5 (Hypothetical protein P0038D11.30).  
 GN Name=OJ1316\_H05.5; Synonyms=P0038D11.30;  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzeae; Oryza.  
 OX NCBI\_TaxID=39947;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Sasaki T., Matsumoto T., Yamamoto K., Sakata K., Baba T., Katayose Y.,  
 RA Wu J., Nimura Y., Cheng Z., Nagamura Y., Antonio B.A., Kanamori H.,  
 RA Hosokawa S., Masukawa M., Arikawa K., Chiden Y., Hayashi M.,  
 RA Okamoto M., Ando T., Aoki H., Arita K., Hamada M., Harada C.,  
 RA Hijishita S., Honda M., Ichikawa Y., Idonuma A., Iijima M., Ikeda M.,  
 RA Ikeno M., Itoh S., Itoh Y., Itoh Y., Iwabuchi A., Kamiya K.,  
 RA Karasawa W., Katagiri S., Kikuta A., Kobayashi N., Kono I.,  
 RA Machita K., Maehara T., Mizuno H., Mizubayashi T., Mukai Y.,  
 RA Nagasaki H., Nakashima M., Nakama Y., Nakamichi Y., Nakamura M.,  
 RA Namiki N., Negishi M., Ohta I., Ono N., Saji S., Sakai K., Shibata M.,  
 RA Shimokawa T., Shomura A., Song J., Takazaki Y., Terasawa K., Tsuji K.,  
 RA Waki K., Yamagata H., Yamane H., Yoshiki S., Yoshihara R., Yukawa K.,  
 RA Zhong H., Iwama H., Endo T., Ito H., Hahn J.H., Kim H.I., Eun M.Y.,  
 RA Yano M., Jiang J., Gojobori T.;  
 RT "The genome sequence and structure of rice chromosome 1.";  
 RL Nature 420:312-316(2002).  
 DR EMBL; AP003340; BAD68324.1; -; Genomic DNA.  
 DR EMBL; AP003234; BAD68191.1; -; Genomic DNA.  
 KW Hypothetical protein.  
 SQ SEQUENCE 76 AA; 7923 MW; 2569B99B4D597000 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 76;

Best Local Similarity 100.0%; Pred. No. 3.7e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5

Db 50 RGRGR 54

```

RESULT 26
ID Q6K8Z0_ORYSA PRELIMINARY; PRT; 76 AA.
AC Q6K8Z0_
DT 05-JUL-2004 (T-EMBLrel. 27, Created)
DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)
DE Hypothetical protein QJ1695_H09.12.
GN Name=QJ1695_H09.12;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF004094; BAD19270.1; -; Genomic_DNA.
DR Gramene; Q6K8Z0; -.
KW Hypothetical protein.
SQ SEQUENCE 76 AA; 8106 MW; 66488ED35E48B721 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 76;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
DB 3 RGRGR 7

RESULT 27
ID Q4SAZ2_TETNG PRELIMINARY; PRT; 77 AA.
AC Q4SAZ2;
DT 13-SEP-2005 (T-EMBLrel. 31, Created)
DT 13-SEP-2005 (T-EMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (T-EMBLrel. 31, Last annotation update)
DE Chromosome undetermined SCAF14678, whole genome shotgun sequence.
GN ORFNames=GSTENG00021198001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
RA Maccall E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA Blemond C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA Craud C., Duprat S., Brottier P., Couranceau J.P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissenbach J., Roest Crolius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
preliminary data.

```

```

DR EMBL; CAAE01014678; CAG02190.1; -; Genomic DNA.
SQ SEQUENCE 77 AA; 9119 MW; 08B3D48C69B53873 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 77;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
DB 67 RGRGR 71

RESULT 28
ID Q7F7E9_ORYSA PRELIMINARY; PRT; 79 AA.
AC Q7F7E9;
DT 10-MAY-2005 (T-EMBLrel. 30, Created)
DT 10-MAY-2005 (T-EMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (T-EMBLrel. 30, Last annotation update)
DE OSJNBa0036S02.2 protein.
GN Names=OSJNBa0036S02.2;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF002862; BAB17728.1; -; Genomic_DNA.
DR Gramene; Q9SD15; -.
SQ SEQUENCE 79 AA; 8556 MW; 1BBF05E721B69275 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 79;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
DB 3 RGRGR 7

RESULT 29
ID Q8V7D0_9VIRU PRELIMINARY; PRT; 79 AA.
AC Q8V7D0_
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE ORF1 (Fragment).
GN Name=ORF1;
OS Torque teno virus.
OC Viruses; ssDNA viruses; Anellovirus.
OX NCBI_TaxID=68887;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21844401; PubMed=11855633; DOI=10.1007/s705-002-8301-7;
RA Peng Y.H., Nishizawa T., Takahashi M., Ishikawa T., Yoshikawa A.,
RA Okamoto H.;
RT "Analysis of the complete genomes of thirteen TT virus variants
RT classifiable into the fourth and fifth genetic groups, isolated from
RT viremic infants.";
RL Arch. Virol. 147:21-41(2002).
DR EMBL; AB064623; BAB79390.1; -; Genomic_DNA.
FT NON_TER 79
SQ SEQUENCE 79 AA; 10933 MW; 5F0AE2933A4C1A06 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 79;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5

```

```

Db          31 RGRGR 35
|||||
RESULT 30
Q8V7E5_9VIRU PRELIMINARY;      PRT;      79 AA.
AC Q8V7E5_9VIRU PRELIMINARY;      PRT;      79 AA.
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE ORF1 (Fragment).
GN Name=ORF1;
OS Torque teno virus.
OC Viruses; ssDNA viruses; Anellovirus.
OX NCBI_TaxID=68887;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=2184401; PubMed=11855633; DOI=10.1007/s705-002-8301-7;
RA Peng Y.H., Nishizawa T., Takahashi M., Ishikawa T., Yoshikawa A.,
RA Okamoto H.;
RT "Analysis of the complete genomes of thirteen TT virus variants
RT classifiable into the fourth and fifth genetic groups, isolated from
RT viremic infants.";
RL Arch. Virol. 147:21-41(2002).
DR EMBL; AB064613; BAB79370.1; -; Genomic_DNA.
FT NON TER 79
SQ SEQUENCE 79 AA; 10958 MW; FEABEE872E4C0B02 CRC64;

Query Match          100.0%; Score 27; DB 2; Length 79;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          1 RGRGR 5
|||||
Db          31 RGRGR 35
|||||

RESULT 31
Q8LQ79_ORYSA PRELIMINARY;      PRT;      80 AA.
AC Q8LQ79_ORYSA PRELIMINARY;      PRT;      80 AA.
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Hypothetical protein P0702H08.2.
GN Name=P0702H08.2;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX Sasaki T., Matsumoto T., Yamamoto K., Sakata K., Baba T., Katayose Y.,
RA Wu J., Nilamur Y., Cheng Z., Nagamura Y., Antonio B.A., Kanamori H.,
RA Hosokawa S., Masukawa M., Arikawa K., Chiden Y., Hayashi M.,
RA Okamoto M., Ando T., Aoki H., Arita K., Hamada M., Harada C.,
RA Hijishita S., Honda M., Ichikawa Y., Idonuma A., Iijima M., Ikeda M.,
RA Ikeno M., Itoh S., Itoh T., Itoh Y., Itoh Y., Iwabuchi A., Kamiya K.,
RA Karasawa W., Katagiri S., Kikuta A., Kobayashi N., Kono I.,
RA Machita K., Maehara T., Mizuno H., Mizubayashi T., Mukai Y.,
RA Nagasaki K., Nakashima M., Nakama Y., Nakamichi Y., Nakamura M.,
RA Namiki N., Negishi M., Ohta I., Ono N., Saji S., Sakai K., Shibata M.,
RA Shimokawa T., Shomura A., Song J., Takazaki Y., Terasawa K., Tsuji K.,
RA Waki K., Yamagata H., Yamane H., Yoshiki S., Yoshihara R., Yukawa K.,
RA Zhong H., Iwama H., Endo T., Ito H., Hahn J.H., Kim H.I., Eun M.Y.,
RA Yano M., Jiang J., Gojobori T.;
RT "The genome sequence and structure of rice chromosome 1.";
RL Nature 420:312-316(2002).
DR EMBL; AP003350; BAB91848.1; -; Genomic_DNA.
DR Granene; Q8LQ79; -
KW Hypothetical protein.

```

```

SQ SEQUENCE 80 AA; 8343 MW; B21AFDE581BAA4FC CRC64;

Query Match          100.0%; Score 27; DB 2; Length 80;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          1 RGRGR 5
|||||
Db          43 RGRGR 47
|||||

RESULT 32
Q29194_PIG PRELIMINARY;      PRT;      82 AA.
AC Q29194_PIG PRELIMINARY;      PRT;      82 AA.
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Ribosomal protein S2 (Fragment).
OS Sus scrofa (Fig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
OC Sus.
OX NCBI_TaxID=9823;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX TISSUE=Small intestine;
RX MEDLINE=96327607; PubMed=8672129; DOI=10.1007/s003359900153;
RA Winteroe A.K., Fredholm M., Davies W.;
RT "Evaluation and characterization of a porcine small intestine cDNA
RT library: analysis of 839 clones.";
RL Mamm. Genome 7:509-517(1996).
DR EMBL; F14512; CAA23098.1; -; mRNA.
DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
KW Ribosomal protein.
FT NON TER 82
SQ SEQUENCE 82 AA; 9102 MW; 2AA823D3C7543194 CRC64;

Query Match          100.0%; Score 27; DB 2; Length 82;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          1 RGRGR 5
|||||
Db          20 RGRGR 24
|||||

RESULT 33
Q69144_9GAMA PRELIMINARY;      PRT;      82 AA.
AC Q69144_9GAMA PRELIMINARY;      PRT;      82 AA.
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Hypothetical protein (Fragment).
OS Human herpesvirus 4 (Epstein-Barr virus).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10376;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX STRAIN=B95-8;
RX MEDLINE=86094214; PubMed=3001694;
RA Speck S.H., Strominger J.L.;
RT "Analysis of the transcript encoding the latent Epstein-Barr virus
RT nuclear antigen 1: a potentially polycistronic message generated by
RT long-range splicing of several exons.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:8305-8309(1985).
DR EMBL; M12553; AAA6540.1; -; mRNA.
KW Hypothetical protein.
FT NON TER 82
SQ SEQUENCE 82 AA; 8085 MW; F14CB9ABE1FDC09D CRC64;

```



Query Match 100.0%; Score 27; DB 2; Length 82;  
Best Local Similarity 100.0%; Pred. No. 4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
Db 41 RGRGR 45

RESULT 34  
Q5W754 ORYSA  
ID Q5W754\_ORYSA PRELIMINARY; PRT; 83 AA.  
AC Q5W754\_ORYSA PRELIMINARY; PRT; 83 AA.  
DT 01-FEB-2005 (TrEMBLrel. 29, Created)  
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)  
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)  
DE Hypothetical protein QJ1675\_H07.7.  
GN Name=QJ1675\_H07.7;  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=39947;  
RN [1]\_NUCLEOTIDE SEQUENCE.  
RP NUCLEOTIDE SEQUENCE.  
RA Chow T.-Y., Hsing Y.-I.C., Chen C.-S., Chen H.-H., Liu S.-M.,  
RA Chao Y.-T., Chang S.-J., Chen H.-C., Chen S.-K., Chen T.-R.,  
RA Chen Y.-L., Cheng C.-H., Chung C.-I., Han S.-Y., Hsiao S.-H.,  
RA Hsiung J.-N., Hsu C.-H., Huang J.-J., Kau P.-I., Lee M.-C.,  
RA Li Y.-F., Lin S.-J., Lin Y.-C., Wu S.-W., Yu C.-Y., Yu S.-W.,  
RA Wu H.-P., Shaw J.-F., Yu Y., Rambo T., Currie J., Collura K.,  
RA Soderlund C., Wing R.;  
RT "Oryza sativa BAC QJ1675\_H07 genomic sequence."  
RL Submitted (NOV-2004) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC105320; AAV43901.1; -; Genomic\_DNA.  
KW Hypothetical protein.  
SQ SEQUENCE 83 AA; 8745 MW; 199DF3140720185A CRC64;

Query Match 100.0%; Score 27; DB 2; Length 83;  
Best Local Similarity 100.0%; Pred. No. 4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
Db 8 RGRGR 12

RESULT 35  
Q8GVG2\_ORYSA  
ID Q8GVG2\_ORYSA PRELIMINARY; PRT; 83 AA.  
AC Q8GVG2;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)  
DE Hypothetical protein QJ1340\_C08.120 (Hypothetical protein QJ1340\_C08.123).  
GN Name=QJ1340\_C08.120; Synonyms=QJ1340\_C08.123;  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=39947;  
RN [1]\_NUCLEOTIDE SEQUENCE.  
RP NUCLEOTIDE SEQUENCE.  
RA Sasaki T., Matsumoto T., Katayose Y.;  
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, BAC clone:QJ1340\_C08."  
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AP005292; BAC45205.1; -; Genomic\_DNA.  
DR EMBL; AP005292; BAC84241.1; -; Genomic\_DNA.  
DR Gramene; Q8GVG2;  
KW Hypothetical protein.  
SQ SEQUENCE 83 AA; 9206 MW; BA47034C56CAC59 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 83;  
Best Local Similarity 100.0%; Pred. No. 4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
Db 57 RGRGR 61

RESULT 36  
Q67RR9\_SYNTH  
ID Q67RR9\_SYNTH PRELIMINARY; PRT; 83 AA.  
AC Q67RR9;  
DT 25-OCT-2004 (TrEMBLrel. 28, Created)  
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)  
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
DE Hypothetical protein.  
GN OrderedLocustNames=STH639;  
OS Symbiobacterium thermophilum.  
OC Bacteria; Actinobacteria; Symbiobacterium.  
OX NCBI\_TaxID=2734;  
RN [1]\_NUCLEOTIDE SEQUENCE.  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=IAM44863;  
RX PubMed=15383646; DOI=10.1093/nar/gkh830;  
RA Ueda K., Yamashita A., Ishikawa J., Shimada M., Watsui T.,  
RA Morimura K., Ikeda H., Hattori M., Beppu T.;  
RT "Genome sequence of Symbiobacterium thermophilum, an uncultivable bacterium that depends on microbial commensalism."  
RL Nucleic Acids Res. 32:4937-4944 (2004).  
DR EMBL; AP006840; BAD39624.1; -; Genomic\_DNA.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 83 AA; 10301 MW; 662AA7F58260CEA9 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 83;  
Best Local Similarity 100.0%; Pred. No. 4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
Db 23 RGRGR 27

RESULT 37  
Q6H6S6\_ORYSA  
ID Q6H6S6\_ORYSA PRELIMINARY; PRT; 84 AA.  
AC Q6H6S6;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)  
DE Hypothetical protein P0048B08.13 (Hypothetical protein P0519A12.44).  
GN Name=P0048B08.13; Synonyms=P0519A12.44;  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=39947;  
RN [1]\_NUCLEOTIDE SEQUENCE.  
RP NUCLEOTIDE SEQUENCE.  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 2, PAC clone:P0048B08."  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
RN [2]\_NUCLEOTIDE SEQUENCE.  
RP NUCLEOTIDE SEQUENCE.  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 2, PAC clone:P0519A12."  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AP004868; BAD25573.1; -; Genomic\_DNA.  
DR EMBL; AP004839; BAD25573.1; -; Genomic\_DNA.

```
DR Gramene; Q6H656; -.
DR InterPro; IPR005829; Sug transporter.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 84 AA; 8681 MW; 0A1635E815B0F3F8 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 84;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
Db |||||
38 RGRGR 42

RESULT 38
Q6K1R2_ORYSA PRELIMINARY; PRT; 84 AA.
AC Q6K1R2;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein B1215B07.37.
GN Name=B1215B07.37;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Katayose Y.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP006523; BAD23790.1; -; Genomic_DNA.
DR Gramene; Q6K1R2; -.
KW Hypothetical protein.
SQ SEQUENCE 84 AA; 8460 MW; 6F5162FAD822A945 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 84;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
Db |||||
32 RGRGR 36

RESULT 39
Q6Z2C1_ORYSA PRELIMINARY; PRT; 85 AA.
AC Q6Z2C1;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Hypothetical protein OJ1034_C08.17 (Hypothetical protein B1114E07.29).
GN Name=OJ1034_C08.17; Synonyms=B1114E07.29;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Katayose Y.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005383; BAD01368.1; -; Genomic_DNA.

DR Gramene; Q6H656; -.
DR InterPro; IPR005829; Sug transporter.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 84 AA; 8681 MW; 0A1635E815B0F3F8 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 84;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
Db |||||
38 RGRGR 42

RESULT 38
Q6K1R2_ORYSA PRELIMINARY; PRT; 84 AA.
AC Q6K1R2;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein B1215B07.37.
GN Name=B1215B07.37;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Katayose Y.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP006523; BAD23790.1; -; Genomic_DNA.
DR Gramene; Q6K1R2; -.
KW Hypothetical protein.
SQ SEQUENCE 84 AA; 8460 MW; 6F5162FAD822A945 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 84;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
Db |||||
32 RGRGR 36

RESULT 39
Q6Z2C1_ORYSA PRELIMINARY; PRT; 85 AA.
AC Q6Z2C1;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Hypothetical protein OJ1034_C08.17 (Hypothetical protein B1114E07.29).
GN Name=OJ1034_C08.17; Synonyms=B1114E07.29;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Katayose Y.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005383; BAD01368.1; -; Genomic_DNA.

DR Gramene; Q6H656; -.
DR InterPro; IPR005829; Sug transporter.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 84 AA; 8681 MW; 0A1635E815B0F3F8 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 85;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
Db |||||
57 RGRGR 61

RESULT 40
Q6ZLF3_ORYSA PRELIMINARY; PRT; 85 AA.
AC Q6ZLF3;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein OJ1005_H01.26.
GN Name=OJ1005_H01.26;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003798; BAD08734.1; -; Genomic_DNA.
DR Gramene; Q6ZLF3; -.
KW Hypothetical protein.
SQ SEQUENCE 85 AA; 9190 MW; 8230C425E9B0C939 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 85;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
Db |||||
10 RGRGR 14

RESULT 41
Q8H360_ORYSA PRELIMINARY; PRT; 85 AA.
AC Q8H360;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein OSUNBa0077F02.124.
GN Name=OSUNBa0077F02.124;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Katayose Y.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, BAC clone:OSJNBa0077F02.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005247; BAC20861.1; -; Genomic_DNA.
DR Gramene; Q8H360; -.
KW Hypothetical protein.
SQ SEQUENCE 85 AA; 9235 MW; A29DE0791EE299DE CRC64;

Query Match 100.0%; Score 27; DB 2; Length 85;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
```

```
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 RGRGR 5
Db 44 RGRGR 48

RESULT 42
Q6NIQ5_CORDI
ID Q6NIQ5_CORDI PRELIMINARY; PRT; 86 AA.
AC Q6NIQ5_
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Putative transcriptional regulator.
GN OrderedLocusNames=DIP0712;
OS Corynebacterium diphtheriae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1717;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Biotype gravis / NCTC 13129;
RX MEDLINE=22965443; PubMed=14602910; DOI=10.1093/nar/gkg874;
RA Cerdeno-Tarraga A.-M., Estratiou A., Dover L.G., Holden M.T.G.,
RA Pallen M.J., Bentley S.D., Besra G.S., Churcher C.M., James K.D.,
RA De Zoysa A., Chillingworth T., Cronin A., Dowd L., Feltwell T.,
RA Hamlin N., Holtroyd S., Jagella K., Moule S., Quail M.A.,
RA Rabinowitch E., Rutherford K.M., Thomson N.R., Unwin L.,
RA Whitehead S., Barrell B.G., Parkhill J.;
RT "The complete genome sequence and analysis of Corynebacterium
RT diphtheriae NCTC13129.";
RL Nucleic Acids Res. 31:6516-6523 (2003).
DR EMBL; BX248355; CAF49230.1; -; Genomic_DNA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR003482; Whib.
DR Pfam; PF02467; Whib; 1.
KW Complete proteome.
SQ SEQUENCE 86 AA; 9456 MW; 6AA8A65B63B101C CRC64;

Query Match 100.0%; Score 27; DB 2; Length 86;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 RGRGR 5
Db 77 RGRGR 81

RESULT 43
Q79V10_CORGL
ID Q79V10_CORGL PRELIMINARY; PRT; 86 AA.
AC Q79V10; Q8RQC6;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein Cg10768 (transcription factor Whib) (Putative
DE transcriptional regulator Whib).
GN Name=whiB1; Synonyms=whib; OrderedLocusNames=Cg10768; cg0878;
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Nakagawa S.;
RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
```

```
RX PubMed=12948626; DOI=10.1016/S0168-1656(03)00154-8;
RA Kallinowski J., Bathe B., Bartels D., Bischoff N., Bott M.,
RA Burkovski A., Dusch N., Eggeling L., Eikmanns B.J., Gaigalat L.,
RA Goessmann A., Hartmann M., Hutmacher K., Kraemer R., Linke B.,
RA McHardy A.C., Meyer F., Moeckel B., Pfeifferle W., Puhler A.,
RA Rey D.A., Rueckert C., Rupp O., Sahn H., Wendisch V.F., Wiegraebe I.,
RA Tauch A.;
RT "The complete Corynebacterium glutamicum ATCC 13032 genome sequence
RT and its impact on the production of L-aspartate-derived amino acids
RT and vitamins.";
RL J. Biotechnol. 104:5-25 (2003).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RA Kim H.-J., Lee H.-S.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; BA000036; BAB98161.1; -; Genomic_DNA.
DR EMBL; BX927150; CAF19473.1; -; Genomic_DNA.
DR EMBL; AF216692; AAL84640.1; -; Genomic_DNA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR003482; Whib.
DR Pfam; PF02467; Whib; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 86 AA; 9586 MW; ED1FBB572D7E0613 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 86;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 RGRGR 5
Db 77 RGRGR 81

RESULT 44
Q8FRH8_COREF
ID Q8FRH8_COREF PRELIMINARY; PRT; 86 AA.
AC Q8FRH8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Putative transcription regulator.
DE OrderedLocusNames=CE0783;
GN Corynebacterium efficiens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=152794;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RX MEDLINE=22723752; PubMed=12840036; DOI=10.1101/gr.1285603;
RA Nishio Y., Nakamura Y., Kawarabayasi Y., Usuda Y., Kimura E.,
RA Sugimoto S., Matsui K., Yamagishi A., Kikuchi H., Ikeo K.,
RA Gojobori T.;
RT "Comparative complete genome sequence analysis of the amino acid
RT replacements responsible for the thermostability of Corynebacterium
RT efficiens.";
RL Genome Res. 13:1572-1579 (2003).
DR EMBL; BA000035; BAC17593.1; -; Genomic_DNA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR003482; Whib.
DR Pfam; PF02467; Whib; 1.
KW Complete proteome.
SQ SEQUENCE 86 AA; 9586 MW; 01D9C75D26151EA8 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 86;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 RGRGR 5
Db 77 RGRGR 81
```

```

RESULT 45
Q6ERR9 ORYSA
ID Q6ERR9 ORYSA PRELIMINARY; PRT; 87 AA.
AC Q6ERR9;
DT 25-OCT-2004 (TRENBLrel. 28, Created)
DT 25-OCT-2004 (TRENBLrel. 28, Last sequence update)
DE Hypothetical protein P0701F11.15.
GN Name=P0701F11.15;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005429; BAD28651.1; -; Genomic_DNA.
DR Gramene; O6ERE9; -.
KW Hypothetical protein.
SQ SEQUENCE 87 AA; 9222 MW; 98135372A057480C CRC64;

Query Match 100.0%; Score 27; DB 2; Length 87;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
DB 12 RGRGR 16

RESULT 46
Q6K9E3 ORYSA
ID Q6K9E3 ORYSA PRELIMINARY; PRT; 87 AA.
AC Q6K9E3;
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DE Hypothetical protein OJ1471_E11.14.
GN Name=OJ1471_E11.14;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP004062; BAD22987.1; -; Genomic_DNA.
DR Gramene; Q6K9E3; -.
KW Hypothetical protein.
SQ SEQUENCE 87 AA; 9021 MW; 9AE798EC6D837D5C CRC64;

Query Match 100.0%; Score 27; DB 2; Length 87;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
DB 5 RGRGR 9

RESULT 47
Q6ZAG7 ORYSA
ID Q6ZAG7 ORYSA PRELIMINARY; PRT; 87 AA.
AC Q6ZAG7;
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DT 01-FEB-2005 (TRENBLrel. 29, Last annotation update)

```

```

DE Hypothetical protein P0048G02.1 (Hypothetical protein
DE OJ1117_F10.34).
GN Name=P0048G02.1; Synonyms=OJ1117_F10.34;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP004662; BAD09778.1; -; Genomic_DNA.
DR EMBL; AP003871; BAD08798.1; -; Genomic_DNA.
DR Gramene; Q6ZAG7; -.
KW Hypothetical protein.
SQ SEQUENCE 87 AA; 9024 MW; D94CA206BEB6B83B CRC64;

Query Match 100.0%; Score 27; DB 2; Length 87;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
DB 18 RGRGR 22

RESULT 48
Q6ER89 ORYSA
ID Q6ER89 ORYSA PRELIMINARY; PRT; 88 AA.
AC Q6ER89;
DT 25-OCT-2004 (TRENBLrel. 28, Created)
DT 25-OCT-2004 (TRENBLrel. 28, Last sequence update)
DT 10-MAY-2005 (TRENBLrel. 30, Last annotation update)
DE Epstein-Barr virus EBNA-1-like.
GN Name=OSJNBa0014M17.13; Synonyms=P0508B05.34;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Katayose Y.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 2, PAC
clone:P0508B05."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005609; BAD28831.1; -; Genomic_DNA.
DR EMBL; AP004753; BAD27920.1; -; Genomic_DNA.
DR Gramene; Q6ER89; -.
SQ SEQUENCE 88 AA; 9662 MW; 2CFE388FA6FFB4F9 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 88;
Best Local Similarity 100.0%; Pred. No. 4.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
DB 52 RGRGR 56

RESULT 49
Q6YYW8 ORYSA
ID Q6YYW8 ORYSA PRELIMINARY; PRT; 89 AA.
AC Q6YYW8;

```

```

DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein P0604E01.30.
GN Names=P0604E01.30;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Katayose Y.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005544; BAD13216.1; -; Genomic_DNA.
DR Gramene; O6YVW8; -.
KW Hypothetical protein.
SQ SEQUENCE 89 AA; 9625 MW; FC2BD77F7000BEB2 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 89;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRGR 5
Db 41 RGRGR 45

RESULT 50
Q8LH02_ORYSA
ID Q8LH02_ORYSA PRELIMINARY; PRT; 89 AA.
AC Q8LH02;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein P0446F04.111.
GN Names=P0446F04.111;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Katayose Y.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, PAC
clone:P0446F04.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005187; BAC10395.1; -; Genomic_DNA.
DR Gramene; Q8LH02; -.
KW Hypothetical protein.
SQ SEQUENCE 89 AA; 9920 MW; FB6033AC01B9D2A5 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 89;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRGR 5
Db 15 RGRGR 19

RESULT 51
Q8FQ78_COREF
ID Q8FQ78_COREF PRELIMINARY; PRT; 89 AA.
AC Q8FQ78;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN OrderedLocustNames=CE1255;
OC Corynebacterium efficiens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

```

```

OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=152794;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RX MEDLINE=22723752; PubMed=12840036; DOI=10.1101/gr.1285603;
RA Nishio Y., Nakamura Y., Kawarabayashi Y., Usuda Y., Kimura E.,
RA Sugimoto S., Matsui K., Yamagishi A., Kikuchi H., Ikeo K.,
RA Gojobori T.;
RT "Comparative complete genome sequence analysis of the amino acid
RT replacements responsible for the thermostability of Corynebacterium
RT efficiens.";
RL Genome Res. 13:1572-1579(2003).
DR EMBL; BA000035; BAC18065.1; -; Genomic_DNA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR002145; HTH_CoPG.
DR Pfam; PF01402; RHH_1; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 89 AA; 9592 MW; 0221FDE9511141BA CRC64;

Query Match 100.0%; Score 27; DB 2; Length 89;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRGR 5
Db 38 RGRGR 42

RESULT 52
Q84ZB3_ORYSA
ID Q84ZB3_ORYSA PRELIMINARY; PRT; 90 AA.
AC Q84ZB3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Hypothetical protein P0705A05.108 (Hypothetical protein P0543D10.39)
GN Names=P0705A05.108; Synonyms=P0104B02.9; P0543D10.39;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Katayose Y.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP004623; BAC55992.1; -; Genomic_DNA.
DR EMBL; AP004587; BAD09631.1; -; Genomic_DNA.
DR EMBL; AP006461; BAD10766.1; -; Genomic_DNA.
DR Gramene; Q84ZB3; -.
DR InterPro; IPR000276; GPCR_Rhodopsn.
DR PROSITE; PS00237; G_PROTEIN_RECPT_F1_1; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 90 AA; 9630 MW; B442F3C86C31A05B CRC64;

Query Match 100.0%; Score 27; DB 2; Length 90;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRGR 5
Db 52 RGRGR 56

```

RESULT 53					
Q75150_ORYZA					
ID	Q75150	ORYZA PRELIMINARY;	PRT;	94	AA.
AC	Q75150;				
DT	05-JUL-2004	(TrEMBLrel. 27, Created)			
DT	05-JUL-2004	(TrEMBLrel. 27, Last sequence update)			
DT	05-JUL-2004	(TrEMBLrel. 27, Last annotation update)			
DE	Putative gibberellin regulated protein.				
GN	Name=Os03g941060;				
OS	Oryza sativa (japonica cultivar-group).				
OC	Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;				
OC	Ehrhartoideae; Oryzeae; Oryza.				
OX	NCBI_TaxID=39947;				
RN	[1]				
RP	NUCLEOTIDE SEQUENCE.				
RA	Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,				
RA	Overton II L.L., Tsiirzin T., Kim M.M., Bera J.J., Jin S.,				
RA	Radrosh D.W., Tallon L.J., Koo H., Ziemann V., Hsiao J., Blunt S.,				
RA	Vanaken S.S., Riedmuller S.B., Utterback T.T., Feldblyum T.V.,				
RA	Yang Q., Haas B.J., Suh B.B., Peterson J.J., Quackenbush J., White O.,				
RA	Salzberg S.L., Fraser C.M.;				
RT	"Oryza sativa chromosome 3 BAC OSJNBA0091E13 genomic sequence.";				
RL	Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.				
RN	[2]				
RP	NUCLEOTIDE SEQUENCE.				
RA	Buell R.;				
RL	Submitted (JAN-2005) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AC133860; AAR87222.1; -; Genomic_DNA.				
DR	Gramene; Q75150; -;				
DR	InterPro; IPR003854; GASA.				
DR	Pfam; PF02704; GASA; 1.				
SQ	SEQUENCE 94 AA; 9930 MW; C3FE95FDE3A7F779 CRC64;				
Query Match					
Best Local Similarity 100.0%; Score 27; DB 2; Length 94;					
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	RGRGR 5			
Db	43	RGRGR 47			
RESULT 54					
Q6N9G2_RHOPA					
ID	Q6N9G2	RHOPA PRELIMINARY;	PRT;	94	AA.
AC	Q6N9G2;				
DT	05-JUL-2004	(TrEMBLrel. 27, Created)			
DT	05-JUL-2004	(TrEMBLrel. 27, Last sequence update)			
DT	05-JUL-2004	(TrEMBLrel. 27, Last annotation update)			
DE	Hypothetical protein precursor.				
GN	OrderedLocusNames=RP1587;				
OS	Rhodopseudomonas palustris.				
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;				
OC	Bradyrhizobiaceae; Rhodopseudomonas.				
OX	NCBI_TaxID=1076;				
RN	[1]				
RP	NUCLEOTIDE SEQUENCE.				
RC	STRAIN=CGA009 / ATCC BAA-98;				
RX	PubMed=14704707; DOI=10.1038/nbt923;				
RA	Larimer F.W., Chain P., Hauser L., Lamerdin J.E., Malfatti S., Do L.,				
RA	Land M.L., Pelletier D.A., Beatty J.T., Lang A.S., Tabita F.R.,				
RA	Gibson J.L., Hanson T.E., Bobst C., Torres y Torres J.L., Peres C.,				
RA	Harrison F.H., Gibson J., Harwood C.S.;				
RT	"Complete genome sequence of the metabolically versatile				
RT	photosynthetic bacterium Rhodopseudomonas palustris.";				
RL	Nat. Biotechnol. 22:55-61(2004).				
DR	EMBL; BX572598; CAE27028.1; -; Genomic DNA.				
KW	Complete proteome; Hypothetical protein; Signal.				
FT	SIGNAL 1 34				
SQ	SEQUENCE 94 AA; 9885 MW; ASC0EB71C1777BCF CRC64;				
Query Match					
Best Local Similarity 100.0%; Score 27; DB 2; Length 94;					
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	RGRGR 5			
Db	16	RGRGR 20			
RESULT 55					
Q7NPQ9_CHRVO					
ID	Q7NPQ9	CHRVO PRELIMINARY;	PRT;	94	AA.
AC	Q7NPQ9;				
DT	01-MAR-2004	(TrEMBLrel. 26, Created)			
DT	01-MAR-2004	(TrEMBLrel. 26, Last sequence update)			
DT	01-MAR-2004	(TrEMBLrel. 26, Last annotation update)			
DE	Hypothetical protein.				
GN	OrderedLocusNames=CV4017;				
OS	Chromobacterium violaceum.				
OC	Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;				
OC	Neisseriaceae; Chromobacterium.				
OX	NCBI_TaxID=536;				
RN	[1]				
RP	NUCLEOTIDE SEQUENCE.				
RC	STRAIN=ATCC 12472 / DSM 30191;				
RX	MEDLINE=22882880; PubMed=14500782; DOI=10.1073/pnas.1832124100;				
RA	Vasconcelos A.T.R., de Almeida D.F., Hungria M., Guimaraes C.T.,				
RA	Antonio R.V., Almeida F.C., de Almeida L.G.P., de Almeida R.,				
RA	Alves-Gomes J.A., Andrade E.M., Araripe J., de Araujo M.F.F.,				
RA	Astolfi-Filho S., Azevedo V., Baptista A.J., Bataus L.A.M.,				
RA	Batista J.S., Belo A., van den Berg C., Bogo M., Bonatto S.,				
RA	Bordignon J., Brigido M.M., Brito C.A., Brocchi M., Burity H.A.,				
RA	Camargo A.A., Cardoso D.D.P., Carneiro N.P., Carraro D.M.,				
RA	Carvalho C.M.B., Cascardo J.C.M., Cavada B.S., Chueire L.M.O.,				
RA	Creczynski-Pasa T.B., Cunha-Junior N.C., Fagundes N., Falcao C.L.,				
RA	Fancinatti F., Farias I.P., Felipe M.S.S., Ferrari L.P., Ferro J.A.,				
RA	Ferreto M.I.T., Franco G.R., Freitas N.S.A., Furian L.R.,				
RA	Gazzinelli R.T., Gomes E.A., Goncalves P.R., Grangeiro T.B.,				
RA	Grattapaglia D., Grisard E.C., Hanna E.S., Jardim S.N., Laurino J.,				
RA	Leoi L.C.T., Lima L.F.A., Loureiro M.F., Lyra M.C.C.P.,				
RA	Madeira H.M.F., Manfio G.P., Maranhao A.Q., Martins W.S.,				
RA	di Mauro S.M.Z., de Medeiros S.R.B., Meissner R.V., Moreira M.A.M.,				
RA	Nascimento F.F., Nicolas M.F., Oliveira J.G., Oliveira S.C.,				
RA	Paixao R.F.C., Parente J.A., Pedrosa F.O., Pena S.D.J., Pereira J.O.,				
RA	Pereira M., Pinto L.S.R.C., Pinto L.S., Porto J.I.R., Potrich D.P.,				
RA	Ramalho-Neto C.E., Reis A.M.M., Rigo L.U., Rondinelli E.,				
RA	Santos E.B.P., Santos F.R., Schneider M.P.C., Seunanez H.N.,				
RA	Silva A.M.R., da Silva A.L.C., Silva D.W., Silva R., Simoes I.C.,				
RA	Simon D., Soares C.M.A., Soares R.B.A., Souza E.M., Souza K.R.L.,				
RA	Souza R.C., Steffens M.B.R., Steindel M., Teixeira S.R., Urmenyi T.,				
RA	Vettore A., Wassen R., Zaha A., Simpson A.J.G.;				
RT	"The complete genome sequence of Chromobacterium violaceum reveals				
RT	remarkable and exploitable bacterial adaptability.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 100:11660-11665(2003).				
DR	EMBL; AE016924; AAQ64067.1; -; Genomic DNA.				
KW	Complete proteome; Hypothetical protein.				
SQ	SEQUENCE 94 AA; 10793 MW; 835B4F8072398CE4 CRC64;				

```

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Ferredoxin.
GN OrderedLocusNames=TM1289;
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MSB8 / DSM 3109 / ATCC 43589;
RX MEDLINE=99287316; PubMed=10360571; DOI=10.1038/20601;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L.A., Utterback T.R., Malek J.A., Linner K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.L.,
RA Heidelberg J.F., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
DR EMBL; AE001784; AAD36363.1; -; Genomic_DNA.
DR PIR; H72271; H72271.
DR HSSP; P00195; 1CLF.
DR TIGR; TM1289; -
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001450; 4Fe4S_Fe_S_bd.
DR Pfam; PF00037; Fer4; 2.
DR PRINTS; PR00353; 4FE4SFROXIN.
DR PROSITE; PS00198; 4FE4S_FERREDOXIN; 2.
KW 4Fe-4S; Complete proteome; Electron transport; Iron; Iron-sulfur;
KW Metal-binding; Transport.
SQ SEQUENCE 95 AA; 10304 MW; AFEDB05FF25A34E7 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 95;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRGR 5
Db 87 RGRGR 91

RESULT 57
Q5T6U8 HUMAN
ID Q5T6U8_HUMAN PRELIMINARY; PRT; 96 AA.
AC Q5T6U8;
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE OTTHUMP00000016223 (High mobility group AT-hook 1).
GN Name=HMGAL1; ORFNames=RP11-513I15.2-002;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
RA Williams S.;
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC Kainine N., Chen X., Rolfs A., Halleck A., Hines L., Eisenstein S.,
RA Koundinya M., Raphael J., Moreira D., Kelley T., Labaer J., Lin Y.,
RA Phelan M., Farmer A.;
RT "Cloning of human full-length CDSs in BD Creator(TM) System Donor
RT vector.";
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; ALJ354740; CA114992.1; -; Genomic_DNA.
DR EMBL; BT006774; AAP35420.1; -; mRNA.

```

```

DR GO; GO:0000785; C:chromatin; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0007001; P:chromosome organization and biogenesis (sen...; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
SQ SEQUENCE 96 AA; 10679 MW; B82DCAA29E6D18FD CRC64;

Query Match 100.0%; Score 27; DB 2; Length 96;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRGR 5
Db 24 RGRGR 28

RESULT 58
Q6UQF0 CANFA
ID Q6UQF0_CANFA PRELIMINARY; PRT; 96 AA.
AC Q6UQF0;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE High mobility group protein AlB.
GN Name=HMGALB;
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
OC Canis.
OX NCBI_TaxID=9615;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC PubMed=15087128; DOI=10.1016/j.gene.2004.01.009;
RA Murua Escobar H., Soller J.T., Richter A., Meyer B., Winkler S.,
RA Flohr A.M., Nolte I., Bullerdiek J.;
RT "The canine HMGALB.";
RL Gene 330:93-99(2004).
DR EMBL; AY364003; AARI3046.1; -; mRNA.
DR Ensembl; ENSCARG00000001211; Canis familiaris.
DR GO; GO:0000785; C:chromatin; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0007001; P:chromosome organization and biogenesis (sen...; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000637; A+T hook.
DR InterPro; IPR000116; Highmobility_IY.
DR Pfam; PF02178; AT hook; 3.
DR PRINTS; PR00929; ATHOOK.
DR PRINTS; PR00930; HIGHMOBILITYIY.
DR SMART; SM00384; AT hook; 3.
DR PROSITE; PS00354; HMG1 Y; 3.
SQ SEQUENCE 96 AA; 10649 MW; A59B11D29E6D18FD CRC64;

Query Match 100.0%; Score 27; DB 2; Length 96;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RGRGR 5
Db 24 RGRGR 28

RESULT 59
Q6UQF9 CANFA
ID Q6UQF9_CANFA PRELIMINARY; PRT; 96 AA.
AC Q6UQF9;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE High mobility group protein AlB (High mobility group HMGALB) (High
DE mobility group protein AlB).
GN Name=HMGALB;
OS Canis familiaris (Dog).

```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
OC Canis.
OX NCBI_TaxID=9615;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15087128; DOI=10.1016/j.gene.2004.01.009;
RA Murua Escobar H., Soller J.T., Richter A., Meyer B., Winkler S.,
RT Flohr A.M., Nolte I., Bullerdiek J.;
RL "The canine HMGAL.";
RL Gene 330:93-99(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Murua Escobar H., Soller J.T., Richter A., Meyer B., Nolte I.,
RA Bullerdiek J.;
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY363994; AAR13037.1; -; mRNA.
DR EMBL; AY363995; AAR13038.1; -; mRNA.
DR EMBL; AY363996; AAR13039.1; -; mRNA.
DR EMBL; AY363998; AAR13041.1; -; mRNA.
DR EMBL; AY364000; AAR13043.1; -; mRNA.
DR EMBL; AY364002; AAR13045.1; -; mRNA.
DR EMBL; AY366393; AAR21599.1; -; mRNA.
DR EMBL; AY363611; AAR12008.1; -; mRNA.
DR EMBL; AY366392; AAR21598.1; -; mRNA.
DR EMBL; AY364001; AAR13044.1; -; mRNA.
DR EMBL; AY363999; AAR13042.1; -; mRNA.
DR EMBL; AY363997; AAR13040.1; -; mRNA.
DR Ensembl; ENSCAFG0000001211; Canis familiaris.
DR GO; GO:0000785; C:nucleus; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0007001; P:chromosome organization and biogenesis (sen. . .; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000637; AT hook DNA bd.
DR InterPro; IPR001116; Highmoblty_IV.
DR Pfam; PF02178; AT hook; 3.
DR PRINTS; PR00929; ATHOOK.
DR PRINTS; PR00930; HIGHMOBLTYIY.
DR SMART; SM00384; AT hook; 3.
DR PROSITE; PS00354; HMG1_Y; 3.
SQ SEQUENCE 96 AA; 10679 MW; B82DCAA29E6D18FD CRC64;

Query Match 100.0%; Score 27; DB 2; Length 96;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
Db 24 RGRGR 28

RESULT 60
QGRRS_ORYSA PRELIMINARY; PRT; 96 AA.
AC Q8GRR5;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-FEB-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Hypothetical protein P0492E07.129 (Hypothetical protein
DE P0710F09.104).
GN Name=P0492E07.129; Synonyms=P0710F09.104;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, PAC
RT clone:P0492E07.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

[2]
RN NUCLEOTIDE SEQUENCE.
RP Sasaki T., Matsumoto T., Katayose Y.;
RA "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, PAC
RT clone:P0710F09.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005325; BAD31327.1; -; Genomic_DNA.
DR EMBL; AP004305; BAC20733.1; -; Genomic_DNA.
DR Gramene; Q8GRR5; -.
DR InterPro; IPR000637; A+T hook.
DR PRINTS; PR00929; ATHOOK.
KW Hypothetical protein.
SQ SEQUENCE 96 AA; 10722 MW; C381B193383625C1 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 96;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
Db 51 RGRGR 55

RESULT 61
Q8K1F5 RAT PRELIMINARY; PRT; 96 AA.
AC Q8K1F5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE HMGALb.
GN Name=Hmgal;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Copenhagen; TISSUE=Prostate tumor;
RA Sgarra R., Diana F., Bellarosa C., Rustighi A., Toller M.,
RA Manfioletti G., Giancotti V.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF511040; AAM74157.1; -; mRNA.
DR RGD; 628699; Hmgal.
DR GO; GO:0000785; C:chromatin; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0007001; P:chromosome organization and biogenesis (sen. . .; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000637; A+T hook.
DR InterPro; IPR001116; Highmoblty_IV.
DR Pfam; PF02178; AT hook; 3.
DR PRINTS; PR00929; ATHOOK.
DR PRINTS; PR00930; HIGHMOBLTYIY.
DR ProDom; PD005593; Highmoblty_IV; 1.
DR SMART; SM00384; AT hook; 3.
DR PROSITE; PS00354; HMG1_Y; 3.
SQ SEQUENCE 96 AA; 10709 MW; 1C685D17CB68561D CRC64;

Query Match 100.0%; Score 27; DB 2; Length 96;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
Db 24 RGRGR 28

RESULT 62
Q6W8X3 CHICK PRELIMINARY; PRT; 96 AA.
AC Q6W8X3;
```



DT 05-JUL-2004 (T-EMBLrel. 27, Created)  
 DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)  
 DE High mobility group A1b.  
 GN Name=HMGAL;  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=22812118; PubMed=12930952; DOI=10.1093/nar/gkg684;  
 RA Beitzel B.; Bushman F.;  
 RT "Construction and analysis of cells lacking the HMG A gene family.";  
 RL Nucleic Acids Res. 31:5025-5032(2003).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Beitzel B.F.; Bushman F.D.;  
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY303673; AAQ63840.1; -; mRNA.  
 DR GO; GO:0000785; C:chromatin; IEA.  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR GO; GO:0007001; P:chromosome organization and biogenesis (sen...); IEA.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
 DR InterPro; IPR000637; A+T hook.  
 DR InterPro; IPR001116; Highmobility\_IY.  
 DR Pfam; PF02178; AT hook; 3.  
 DR PRINTS; PR00929; ATHOOK.  
 DR PRINTS; PR00930; HIGHMOBILITYIY.  
 DR ProDom; PD005593; HIGHMOBILITYIY. 1.  
 DR SMART; SM00384; AT hook; 3.  
 DR PROSITE; PS00354; HMG1 Y; 2.  
 SQ SEQUENCE 96 AA; 10429 MW; CA63BADB72289FDD CRC64;  
  
 Query Match 100.0%; Score 27; DB 2; Length 96;  
 Best Local Similarity 100.0%; Pred. No. 4.6e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 1 RGRGR 5  
 DB 24 RGRGR 28  
  
 RESULT 63  
 Q6UF82\_9HIV1  
 ID Q6UF82\_9HIV1 PRELIMINARY; PRT; 96 AA.  
 AC Q6UF82\_9HIV1 PRELIMINARY; PRT; 96 AA.  
 DT 05-JUL-2004 (T-EMBLrel. 27, Created)  
 DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)  
 DE Vpr protein.  
 GN Name=vpr;  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;  
 OC Lentivirus; Primate lentivirus group.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Kijak G.H.; Sanders-Buell E.; Wolfe N.D.; Mpoudi-Ngole E.; Kim B.;  
 RA Robb M.L.; Birx D.L.; Burke D.S.; Carr J.K.; McCutchan F.E.;  
 RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY371142; AAR22072.1; -; Genomic\_DNA.  
 DR HSP; P12520; 1BDE.  
 DR SMR; Q6UF82; 1-96.  
 DR InterPro; IPR000012; Retrov\_Vpr/X.  
 DR Pfam; PF00522; VPR; 1.  
 DR PRINTS; PR00444; HIVVPRPX.  
 DR AIDS.  
 SQ SEQUENCE 96 AA; 11110 MW; 4A25FEB31E5B03C6 CRC64;  
  
 Query Match 100.0%; Score 27; DB 2; Length 96;  
 Best Local Similarity 100.0%; Pred. No. 4.6e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 1 RGRGR 5  
 DB 24 RGRGR 28  
  
 RESULT 64  
 Q6UFA0\_9HIV1  
 ID Q6UFA0\_9HIV1 PRELIMINARY; PRT; 96 AA.  
 AC Q6UFA0\_9HIV1 PRELIMINARY; PRT; 96 AA.  
 DT 05-JUL-2004 (T-EMBLrel. 27, Created)  
 DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)  
 DE Vpr protein.  
 GN Name=vpr;  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;  
 OC Lentivirus; Primate lentivirus group.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Kijak G.H.; Sanders-Buell E.; Wolfe N.D.; Mpoudi-Ngole E.; Kim B.;  
 RA Robb M.L.; Birx D.L.; Burke D.S.; Carr J.K.; McCutchan F.E.;  
 RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY371140; AAR22054.1; -; Genomic\_DNA.  
 DR HSP; P12520; 1BDE.  
 DR SMR; Q6UFA0; 1-96.  
 DR InterPro; IPR000012; Retrov\_Vpr/X.  
 DR Pfam; PF00522; VPR; 1.  
 DR PRINTS; PR00444; HIVVPRPX.  
 DR AIDS.  
 SQ SEQUENCE 96 AA; 11272 MW; 732DE95308DCA484 CRC64;  
  
 Query Match 100.0%; Score 27; DB 2; Length 96;  
 Best Local Similarity 100.0%; Pred. No. 4.6e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 1 RGRGR 5  
 DB 86 RGRGR 90  
  
 RESULT 65  
 Q6ZBK7\_ORYSA  
 ID Q6ZBK7\_ORYSA PRELIMINARY; PRT; 97 AA.  
 AC Q6ZBK7\_ORYSA PRELIMINARY; PRT; 97 AA.  
 DT 05-JUL-2004 (T-EMBLrel. 27, Created)  
 DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)  
 DE Hypothetical protein P0689E12.18.  
 GN Name=P0689E12.18;  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=39947;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Sasaki T.; Matsumoto T.; Yamamoto K.;  
 RA Sasaki T.; Matsumoto T.; Yamamoto K.;  
 RT "Oryza sativa nipponbare (GAJ) genomic DNA, chromosome 8, PAC  
 RT clone: P0689E12.";  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP004622; BAD09692.1; -; Genomic\_DNA.  
 DR Gramene; Q6ZBK7; -;  
 DR InterPro; IPR000637; A+T hook.  
 DR PRINTS; PR000929; ATHOOK.  
 KW Hypothetical protein.  
 SQ SEQUENCE 97 AA; 10782 MW; 1B17D1BB6C461442 CRC64;  
  
 Query Match 100.0%; Score 27; DB 2; Length 97;  
 Best Local Similarity 100.0%; Pred. No. 4.7e+02;

Best Local Similarity 100.0%; Pred. No. 4.6e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 1 RGRGR 5  
 DB 86 RGRGR 90  
  
 RESULT 64  
 Q6UFA0\_9HIV1  
 ID Q6UFA0\_9HIV1 PRELIMINARY; PRT; 96 AA.  
 AC Q6UFA0\_9HIV1 PRELIMINARY; PRT; 96 AA.  
 DT 05-JUL-2004 (T-EMBLrel. 27, Created)  
 DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)  
 DE Vpr protein.  
 GN Name=vpr;  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;  
 OC Lentivirus; Primate lentivirus group.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Kijak G.H.; Sanders-Buell E.; Wolfe N.D.; Mpoudi-Ngole E.; Kim B.;  
 RA Robb M.L.; Birx D.L.; Burke D.S.; Carr J.K.; McCutchan F.E.;  
 RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY371140; AAR22054.1; -; Genomic\_DNA.  
 DR HSP; P12520; 1BDE.  
 DR SMR; Q6UFA0; 1-96.  
 DR InterPro; IPR000012; Retrov\_Vpr/X.  
 DR Pfam; PF00522; VPR; 1.  
 DR PRINTS; PR00444; HIVVPRPX.  
 DR AIDS.  
 SQ SEQUENCE 96 AA; 11272 MW; 732DE95308DCA484 CRC64;  
  
 Query Match 100.0%; Score 27; DB 2; Length 96;  
 Best Local Similarity 100.0%; Pred. No. 4.6e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 1 RGRGR 5  
 DB 86 RGRGR 90  
  
 RESULT 65  
 Q6ZBK7\_ORYSA  
 ID Q6ZBK7\_ORYSA PRELIMINARY; PRT; 97 AA.  
 AC Q6ZBK7\_ORYSA PRELIMINARY; PRT; 97 AA.  
 DT 05-JUL-2004 (T-EMBLrel. 27, Created)  
 DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)  
 DE Hypothetical protein P0689E12.18.  
 GN Name=P0689E12.18;  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=39947;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Sasaki T.; Matsumoto T.; Yamamoto K.;  
 RA Sasaki T.; Matsumoto T.; Yamamoto K.;  
 RT "Oryza sativa nipponbare (GAJ) genomic DNA, chromosome 8, PAC  
 RT clone: P0689E12.";  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP004622; BAD09692.1; -; Genomic\_DNA.  
 DR Gramene; Q6ZBK7; -;  
 DR InterPro; IPR000637; A+T hook.  
 DR PRINTS; PR000929; ATHOOK.  
 KW Hypothetical protein.  
 SQ SEQUENCE 97 AA; 10782 MW; 1B17D1BB6C461442 CRC64;  
  
 Query Match 100.0%; Score 27; DB 2; Length 97;  
 Best Local Similarity 100.0%; Pred. No. 4.7e+02;

```

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 RGRGR 5
Db 92 RGRGR 96

RESULT 66
Q6N2S4 RHOPA PRELIMINARY; PRT; 97 AA.
AC Q6N2S4;
DT 05-JUL-2004 (TREMBLrel. 27, Created)
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DE Hypothetical protein precursor.
GN OrderedLocusNames=RP3A3975;
OS Rhodopseudomonas palustris.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Rhodopseudomonas.
OX NCBI_TaxID=1076;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CGA009 / ATCC BAA-98;
RX PubMed=14704707; DOI=10.1038/nbt923;
RA Larimer F.W., Chain P., Hauser L., Lamerdin J.E., Malfatti S., Do L.,
RA Land M.L., Pelletier D.A., Beatty J.T., Lang A.S., Tabita F.R.,
RA Gibson J.L., Hanson T.E., Bobst C., Torres y Torres J.L., Peres C.,
RA Harrison F.H., Gibson J., Harwood C.S.;
RT "Complete genome sequence of the metabolically versatile
RT photosynthetic bacterium Rhodopseudomonas palustris.";
RL Nat. Biotechnol. 22:55-61(2004).
DR EMBL; BX572605; CAE29416.1; -; Genomic DNA.
KW Complete proteome; Hypothetical protein; Signal.
FT SIGNAL 1 45 Potential.
SQ SEQUENCE 97 AA; 10576 MW; C43BF4532874A3D CRC64;

Query Match 100.0%; Score 27; DB 2; Length 97;
Best Local Similarity 100.0%; Pred. No. 4.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 RGRGR 5
Db 49 RGRGR 53

RESULT 67
Q91148 9HIV1 PRELIMINARY; PRT; 97 AA.
AC Q91148;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Vpr protein.
GN Name=vpr;
OS Human immunodeficiency virus 1.
OC Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;
OC Lentivirus; Primate lentivirus group.
OX NCBI_TaxID=11676;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=BCF03;
RA Bibollet-Ruche F., Lousserat-Ajaka I., Simon F., Mboup S., Mpoudi N.E.,
RA Saman E., Delaporte E., Peeters M.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y16022; CAA75950.1; -; Genomic DNA.
DR HSSP; P12520; 1BDE.
DR SMR; Q91148; 1-96.
DR InterPro; IPR000012; Retrov_Vpr/X.
DR Pfam; PF00522; VPR; 1.
DR PRINTS; PR00444; HIVVPRVPX.
KW AIDS.
SQ SEQUENCE 97 AA; 11266 MW; 710EC232FF6ED56 CRC64;

```

```

Query Match 100.0%; Score 27; DB 2; Length 97;
Best Local Similarity 100.0%; Pred. No. 4.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 RGRGR 5
Db 86 RGRGR 90

RESULT 68
Q77375 9HIV1 PRELIMINARY; PRT; 97 AA.
AC Q77375;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Vpr polyprotein.
GN Name=vpr;
OS Human immunodeficiency virus 1.
OC Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;
OC Lentivirus; Primate lentivirus group.
OX NCBI_TaxID=11676;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=94149849; PubMed=8107220;
RA Vanden Haesevelde M., Decourt J.L., De Leys R.J., Vanderborght B.,
RA van der Groen G., van Heuverswijn H., Saman E.;
RT "Genomic cloning and complete sequence analysis of a highly divergent
RT African human immunodeficiency virus isolate.";
RL J. Virol. 68:1586-1596(1994).
DR EMBL; L20587; AAA99881.1; -; Genomic RNA.
DR HSSP; P05954; 1FI0.
DR SMR; Q77375; 1-96.
DR InterPro; IPR000012; Retrov_Vpr/X.
DR Pfam; PF00522; VPR; 1.
DR PRINTS; PR00444; HIVVPRVPX.
KW AIDS; Polyprotein.
SQ SEQUENCE 97 AA; 11259 MW; 0E118346DB4B4EA7 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 97;
Best Local Similarity 100.0%; Pred. No. 4.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 RGRGR 5
Db 86 RGRGR 90

RESULT 69
Q99D95 9HIV1 PRELIMINARY; PRT; 97 AA.
AC Q99D95;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Vpr protein.
GN Name=vpr;
OS Human immunodeficiency virus 1.
OC Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;
OC Lentivirus; Primate lentivirus group.
OX NCBI_TaxID=11676;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Vallejo A., Gurtler L., Zekeng L., Hewlett I.K.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF316862; AAKJ4221.1; -; Genomic DNA.
DR HSSP; P12520; 1BDE.
DR SMR; Q99D95; 1-96.
DR InterPro; IPR000012; Retrov_Vpr/X.
DR Pfam; PF00522; VPR; 1.
DR PRINTS; PR00444; HIVVPRVPX.
KW AIDS.
SQ SEQUENCE 97 AA; 11287 MW; 0C912FB6344AD755 CRC64;

```

Query Match 100.0%; Score 27; DB 2; Length 97;  
 Best Local Similarity 100.0%; Pred. No. 4.7e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
 Db 86 RGRGR 90  
 |||||

## RESULT 70

Q9WQ16\_9HIV1 PRELIMINARY; PRT; 97 AA.  
 AC Q9WQ16\_9HIV1 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Vpr.  
 GN Names=vpr;  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;  
 OC Lentivirus; Primate lentivirus group.  
 OC NCBI\_TaxID=11676;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=GR84;  
 RX MEDLINE=99284362; PubMed=10357470; DOI=10.1089/088922299310836;  
 RA Nasioulas G., Paraskevis D., Magiorkinis E., Theodoridou M.,  
 RA Hatzakis A.;  
 RT "Molecular analysis of the full-length genome of HIV type 1 subtype I:  
 evidence of A/GI recombination.";  
 RL AIDS Res. Hum. Retroviruses 15:745-758 (1999).  
 DR EMBL; AF119819; AAD38884.1; -; Genomic\_DNA.  
 DR HSSP; P12520; 1BDE.  
 DR SMR; Q9WQ16; 1-92.  
 DR InterPro; IPR000012; Retrov\_Vpr/X.  
 DR Pfam; PF00522; VPR; 1  
 DR PRINTS; PR00444; HIVPRVFX.  
 DR AIDS.  
 KW AIDS.  
 SQ SEQUENCE 97 AA; 11592 MW; 01537E0295754263 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 97;  
 Best Local Similarity 100.0%; Pred. No. 4.7e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
 Db 89 RGRGR 93  
 |||||

## RESULT 71

Q522E5\_NOCFA PRELIMINARY; PRT; 98 AA.  
 AC Q522E5;  
 DT 25-OCT-2004 (TREMBlrel. 28, Created)  
 DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)  
 DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)  
 DE Hypothetical protein.  
 GN OrderedLocusNames=nfa5510;  
 OS Nocardia farcinica  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Nocardiaceae; Nocardia.  
 OC NCBI\_TaxID=37329;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=IFM 10152;  
 RX PubMed=15466710; DOI=10.1073/pnas.0406410101;  
 RA Ishikawa J., Yamashita A., Mikami Y., Hoshino Y., Kurita H., Hotta K.,  
 RA Shiba T., Hattori M.;  
 RT "The complete genomic sequence of Nocardia farcinica IFM 10152.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 101:14925-14930 (2004).  
 DR EMBL; AP006618; BAD55396.1; -; Genomic\_DNA.  
 KW Complete proteome; Hypothetical protein.  
 SQ SEQUENCE 98 AA; 10680 MW; 659E81B8047BABA2 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 98;  
 Best Local Similarity 100.0%; Pred. No. 4.7e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
 Db 18 RGRGR 22  
 |||||

## RESULT 72

Q9FBW0\_STRCO PRELIMINARY; PRT; 98 AA.  
 AC Q9FBW0;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Hypothetical protein SCO7825  
 GN OrderedLocusNames=SCO7825; ORFNames=SC8E7.22;  
 OS Streptomyces coelicolor.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomyces.  
 OC NCBI\_TaxID=1902;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=A3(2) / M145;  
 RX MEDLINE=21996410; PubMed=12000953; DOI=10.1038/417141a;  
 RA Bentley S.D., Chater K.F., Cardeno-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Frazer A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kieser T., Larke L., Murphy L.D., Oliver K., O'Neill S.,  
 RA Rabinowitz E., Rajandream M.A., Rutherford K.M., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J.R., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2).";  
 RL Nature 417:141-147 (2002).  
 DR EMBL; AL939132; CAC03643.1; -; Genomic DNA.  
 KW Complete proteome; Hypothetical protein.  
 SQ SEQUENCE 98 AA; 10069 MW; 447DA028418A6DB3 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 98;  
 Best Local Similarity 100.0%; Pred. No. 4.7e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5  
 Db 76 RGRGR 80  
 |||||

## RESULT 73

Q6JNA8\_9HIV1 PRELIMINARY; PRT; 98 AA.  
 AC Q6JNA8;  
 DT 05-JUL-2004 (TREMBlrel. 27, Created)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
 DE Vpr protein.  
 GN Name=vpr;  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;  
 OC Lentivirus; Primate lentivirus group.  
 OC NCBI\_TaxID=11676;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX PubMed=15186520; DOI=10.1089/088922204323087705;  
 RA Tovanabutra S., Beyrer C., Sakthachornphop S., Razak M.H., Ramos G.L.,  
 RA Tongachak T., Rungruengthanakit K., Sackthio P., Tejafoong K., Kim B.,  
 RA De Souza M., Robb M.L., Birk D.L., Jittiwutikarn J., Suriyanon V.,  
 RA Celentano D.D., McCutchan F.E.;  
 RT "The Changing Molecular Epidemiology of HIV Type 1 among Northern Thai

```

RT Drug Users, 1999 to 2002.;
RL AIDS Res. Hum. Retroviruses 20:465-475(2004).
DR EMBL; AY358069; RAQ98561.1; -; Genomic_DNA.
DR HSP; P12320; 1BDE.
DR SMR; Q6JNAB; 1-96.
DR InterPro; IPR000012; Retrov_Vpr/X.
DR Pfam; PF00522; VPR; 1.
DR PRINTS; PR00444; HIVVPRPX.
KW AIDS.
SQ SEQUENCE 98 AA; 11323 MW; 9E2ABD39097F8B25 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 98;
Best Local Similarity 100.0%; Pred. No. 4.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
DB 88 RGRGR 92

RESULT 74
Q6UFK9_9HIV1
ID Q6UFK9_9HIV1 PRELIMINARY; PRT; 98 AA.
AC Q6UFK9;
DT 05-JUL-2004 (TREMBLrel. 27, Created)
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE Vpr protein.
GN Name=vpr;
OS Human immunodeficiency virus 1.
OC Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae;
OC Lentivirus; Primate lentivirus group.
OX NCBI_TaxID=11676;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Kijak G.H., Sanders-Buell E., Wolfe N.D., Mpoudi-Ngole E., Kim B.,
RA Robb M.L., Birk D.L., Burke D.S., Carr J.K., McCutchan F.E.;
RL Submitted (AUG-2003) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AY371127; RAR21945.1; -; Genomic_DNA.
DR HSP; P12520; 1BDE.
DR SMR; Q6UFK9; 1-96.
DR InterPro; IPR000012; Retrov_Vpr/X.
DR Pfam; PF00522; VPR; 1.
DR PRINTS; PR00444; HIVVPRPX.
KW AIDS.
SQ SEQUENCE 98 AA; 11530 MW; F1F25EF7A501CA02 CRC64;

Query Match 100.0%; Score 27; DB 2; Length 98;
Best Local Similarity 100.0%; Pred. No. 4.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
DB 88 RGRGR 92

RESULT 75
Q8U540_AGR75
ID Q8U540_AGR75 PRELIMINARY; PRT; 99 AA.
AC Q8U540;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE AGR_C_4419P.
GN OrderedLocusNames=AGR_C_4419;
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Cereon;
RX MEDLINE=21608551; PubMed=11743194; DOI=10.1126/science.1066803;

```

```

RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Quorllo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Dougherty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328(2001).
DR EMBL; AE008157; AAK88173.1; -; Genomic_DNA.
DR PIR; D97652; D97652.
SQ SEQUENCE 99 AA; 11444 MW; B7947CF4347F620B CRC64;

Query Match 100.0%; Score 27; DB 2; Length 99;
Best Local Similarity 100.0%; Pred. No. 4.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RGRGR 5
DB 29 RGRGR 33

```

Search completed: December 2, 2005, 10:07:47  
Job time : 181 secs